



Arbeitspapiere und Materialien zur deutschen Sprache

Elke Donalies

Tagtraum, Tageslicht, Tagedieb

**Ein korpuslinguistisches Experiment zu
variierenden Wortformen und Fugenelementen
in zusammengesetzten Substantiven**

**Mit einem Exkurs und zahlreichen Statistiken
von Noah Bubenhofer**



amades

Arbeitspapiere und Materialien zur deutschen Sprache

Herausgegeben vom Institut für Deutsche Sprache

Band 42

1. Auflage 2011 · ISBN: 978-3-937241-34-0 · ISSN: 1435-4195

© 2011 Institut für Deutsche Sprache
R 5, 6-13, 68161 Mannheim
www.amades.de



Mitglied der

Leibniz
Gemeinschaft

Redaktion + Layout: Norbert Volz

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechts ist ohne Zustimmung der Copyright-Inhaber unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Herstellung und Vertrieb im Eigenverlag.

Inhalt

1.	Das Projekt	5
2.	Forschungs(gegen)stand	7
3.	(In)Fragestellung	13
4.	Korpus und Analysemethode	15
Exkurs: Zur automatischen Analyse von Zusammensetzungen		19
5.	Resultate	29
5.1	Regeln.....	33
5.2	Vorkommen und Verteilung von Varianten.....	48
6.	Fazit	103
7.	Literatur	107

1. Das Projekt

Nachdem ein homogener Sprachbegriff in der Linguistik lange dominierte¹ und die Variation aus Gründen der Methode aus der Beschreibung der Sprachsysteme eliminiert wurde, ist die sprachliche Heterogenität nun wiederum als Problem der Sprachwirklichkeit und der adäquaten Sprachbeschreibung besser erkennbar geworden. (Lüdtke/Mattheier 2005, S. 25).

Im empirisch ausgerichteten Projekt „Grammatische Variation im Deutschen“ des IDS wollen wir den Sprachgebrauch in seiner Vielfalt und die tatsächlich wirksamen Regeln darstellen.² Dazu schöpfen wir die heutigen Möglichkeiten einer Korpusgrammatik aus; wir analysieren auf einer möglichst großen Datenbasis grammatische Phänomene mit konkurrierender Varietät.

Meine Studie ist die erste der Pilotstudien, die ein solches Vorgehen ausloten sollen. Dazu hat Noah Bubenhofer ein Versuchskorpus mit 176 405 282 Analysen von Zusammensetzungen aus Substantiven erstellt. Auf Basis dieses Korpus beschreibe ich die sprachgebräuchlichen Varianten von Wortformen und Fugenelementen in deutschen Zusammensetzungen, zum Beispiel in *Tagtraum* neben *Tageslicht* neben *Tagedieb* oder in *Abfahrtsmöglichkeit* neben *Abfahrtnmöglichkeit*. Obwohl solche Varianten immer wieder zu Verwendungsunsicherheiten führen und in der Forschungsliteratur auch immer wieder als auffällig thematisiert werden, ist – wie Michel (2009, S. 334) feststellt – die systematische Beschreibung solcher Varianten ein Desiderat.

Meine Studie richtet sich primär an germanistische Sprachwissenschaftler im In- und Ausland. Durch den besonderen Blick auf die Variation können hier aber auch Didaktiker und andere sprachreflektierende Wissenschaftler und Laien nachschlagen. Schon deshalb habe ich diese Studie allgemeinverständlich gehalten.

¹ Unter anderem weil „die theoretischen Zentralbegriffe der modernen Linguistik, also Saussures langue- und Chomskys Kompetenzbegriff, Homogenität postulieren“ (Schmidt 2005, S. 62).

² Allgemeine Informationen zum Projekt unter www.ids-mannheim.de/gra/korpusgrammatik.html.

Ich danke Gisela Zifonun und Bruno Strecker für anregende Hinweise und Caren Brinckmann für ihre so klare wie konstruktive Kritik. Ulrike Stölzel danke ich für anschauliche Balkendiagramme und Norbert Volz für seine redaktionellen Mühen. Besonders aber danke ich Noah Bubenhofer für die freundliche und effiziente Unterstützung meiner Studie.

Mannheim, August 2011

Elke Donalies

2. Forschungs(gegen)stand

Typischerweise setzen wir im Deutschen Substantive mit Substantiven zusammen, zum Beispiel *Tageslicht*. Dieser typische Typ der Zusammensetzung besteht aus einem Erstglied und einem Zweitglied.

Tageslicht

Erstglied Zweitglied

- Das **Zweitglied**, auch Grundwort oder Determinatum genannt, ist die Hauptsache einer Zusammensetzung. Es entscheidet über die textliche Einbindung der Zusammensetzung, vor allem über Merkmale wie Genus, Kasus und Numerus: *Licht* legt zum Beispiel fest, dass *Tageslicht* ein neutrales Substantiv ist und zu genitivischem *des Tageslichtes* flektiert werden kann. Außerdem entscheidet das Zweitglied über die Bedeutungsmerkmale der Zusammensetzung; es legt die Hauptbedeutung der Zusammensetzung fest: *Tageslicht* bezeichnet in erster Linie ‘Licht’.
- Das **Erstglied**, auch Bestimmungswort oder Determinans genannt, ist dem Zweitglied untergeordnet. Einmal mit dem Zweitglied zusammengesetzt, ist das Erstglied für weitere Operationen unerreichbar. Es wird nicht weiter flektiert; es wird eingefroren. So wird *Apfeltorte* nicht pluralisiert zu *die Äpfeltorten*, sondern zu *die Apfeltorten*. Das Erstglied hat vor allem die Funktion, die Bedeutung des Zweitglieds zu nuancieren: *Tages* nuanciert *Licht* als ein spezielles Licht.

Zusammensetzungen dieses typischen, häufigsten und unbeschränktesten Typs, sogenannte Determinativkomposita, sind in vielen Aspekten für Linguisten und Laien interessant. Mich interessiert hier ausschließlich die Schnittstelle zwischen Erst- und Zweitglied.

Was genau dort passiert, beschreiben Linguisten unterschiedlich. Im Wesentlichen gibt es zwei Positionen.

- **Position 1** geht davon aus, dass Zusammensetzungen aus Substantiven prinzipiell nur aus Stämmen zusammengesetzt werden. Ein Stamm ist eine virtuelle Grundform, ohne Kasus und Numerus, ohne Flexionselemente. Zum Beispiel ist der Stamm von *Tag*, *Tages*, *Tagen* die Grundform *Tag*. Der Stamm, die Grundform stimmt bei Substantiven mit der Nominativ-

Singular-Form überein.³ Alles, was zwischen dem ersten und dem zweiten Stamm steht, wird Fugenelement genannt. Die Verfechter dieser Position sehen das *es* zwischen den Stämmen *Tag* und *Licht* als Fugenelement.

Tag + es + licht

Stamm + **Fugenelement** + Stamm

- **Position 2** geht davon aus, dass das Erstglied eine in realen Texten vorkommende Wortform ist. Zum Beispiel ist *Tages* ein Genitiv Singular. Die Verfechter dieser Position sehen das *es* als das Flexionselement dieser Wortform.

Tages + licht

Wortform + Wortform

Betrachten wir zuerst die äußere Gestalt solcher Zusammensetzungen.

Die sprachgeschichtlich ältesten Zusammensetzungen des Deutschen gehen auf Genitivphrasen wie althochdeutsch *gotes boto* oder *wintes brut* zurück.⁴ In dem sich entwickelnden Zusammensetzungstyp *Gottesbote* – seit Jacob Grimm auch uneigentliche Zusammensetzung genannt – ist das flektierte Erstglied fossilisiert. Analog dazu werden seither Zusammensetzungen gebildet, deren Erstglied teils flektierten Wortformen entspricht, teils aber auch nicht. So kommt *Tages* im Flexionsparadigma von *Tag* vor, dagegen *Kindheits* im Flexionsparadigma von *Kindheit* nicht.

- **Position 1** sieht beide als Formen mit Fugenelement und differenziert in einerseits paradigmatische Fugenelemente, das heißt Fugenelemente, die dem Flexionsparadigma des Erstglieds entsprechen, und andererseits unparadigmatische Fugenelemente. Position 1 nimmt daher eine Vielzahl von Fugenelementen an. Für die einheimische Substantivbildung sind es: „*n* (*Blumenvase*), *s* (*Zweifelsfall*), *ns* (*Glaubensfrage*), *e* (*Pferdewagen*), *er* (*Kindergarten*), *en* (*Heldenmut*), *es* (*Siegeswille*) und *ens* (*Schmerzenschrei*)“ (Eisenberg 2004, S. 236). Kürschner (2003, S. 89) weist darüber hinaus auf Einzelfälle bei Lehnwörtern hin wie *Embryonenforschung*, bei

³ Ausnahmen sind Tilgungsformen wie *Erd* oder *Woll* zum Beispiel in *Erdanziehung* und *Wollknäuel*.

⁴ „Bei dem Zusammenwachsen des Genitivs mit dem regierenden Substantivum im Deutschen ist zunächst zu beachten, dass es nur bei Voranstellung des Genitivs eintritt“ (Paul 1975, S. 335). Vgl. Fleischer/Barz (1995, S. 136), Ramers (1997, S. 36), Eichinger (2000, S. 1), Wegener (2003), Pavlov (2004) und Michel (2009).

dem zum Stamm *Embryo* ein Fugenelement *nen* tritt. Nach Position 1 steht bei Zusammensetzungen mit paradigmatischem Erstglied „die Form des Fugenelements in engem Zusammenhang mit dem Flexionssystem des Erstgliedes“ (Fuhrhop 1998, S. 187). Zum Beispiel heißt es dem Paradigma von *Blume* entsprechend *Blumenvase* und nicht *Blumesvase* oder *Blumervase*.

- **Position 2** hält diese Übereinstimmung der Fugen- mit den Flexionselementen für so wenig zufällig wie die historisch „parallel verlaufenden Grammatikalisierungspfade“ (Wegener 2005, S. 176).⁵ Auch will Position 2 eine Erklärung für den Typ von Erstglied, der dem Flexionsparadigma entsprechend immer umgelautete Formen hat wie in *Ärztelhaus* oder *Bücherregal*.⁶ Niemals kommen hier Formen vor wie *Raderwerk* oder *Haussermeer*. Michel (2009, S. 335) sieht darin eine Wortbildungsspezialität.⁷ Aber was einer Pluralform so exakt entspricht, ist nach Position 2 auch eine Pluralform. Position 2 erschwert zudem die Analyse von Zusammensetzungen wie *Mythenwesen* oder *Firmenwagen*: Position 1 muss dabei ja einen umständlichen Umweg nehmen von der Tilgung eines Elements *os* oder *a* bis zum Ersatz durch ein Fugenelement *en*;⁸ Position 2 stellt einfach fest, dass es sich um die regulären paradigmatischen Pluralformen *Mythen* und *Firmen* handelt. Auch mit *Embryonen* hat Position 2 insofern keine Schwierigkeiten. Position 2 geht ja davon aus, dass in der Wortbildung generell Wortformen verwendet werden.⁹ Danach gibt es bei

⁵ So schon Augst (1975, S. 81). Vgl. dazu auch Kürschner (2003, S. 35-38).

⁶ Der Logik von Position 1 nach müsste es dann auch ein Fugenelement *st* geben in Zusammensetzungen wie *höchstanständig*.

⁷ Schon Fuhrhop (1998) geht von Formen aus, die speziell in der Wortbildung verwendet werden, also offenbar separat in unseren Hirnen abgespeichert sind. Erstglieder in Zusammensetzungen nennt sie „Kompositastammformen“. Warum diese speziellen Formen genauso aussehen wie Flexionsformen, erklärt diese Idee nicht. Wie ist es dazu gekommen, wie kommt es immer noch dazu, dass wir beim Zusammensetzen nicht besondere Wortbildungsformen wählen, sondern Formen, wie wir sie auch aus der Flexion kennen? Und hat schon mal jemand geprüft, ob Nichtlinguisten in Zusammensetzungen wie *Kindergarten* oder *Ärztelhaus* nicht doch einen Plural identifizieren?

⁸ Vgl. unter anderen Ortner et al. (1991, S. 70) oder Fleischer/Barz (1995, S. 136).

⁹ So auch in Zusammensetzungen aus Adjektiven und Substantiven, zum Beispiel in *Bestnote*, *Kleinstwagen* oder *Billigsttourismus*: *Nicht immer freiwillig suchten sie deshalb den Ausweg eher im höherwertigen Reisegeschäft denn im Billigsttourismus mit seinen dürftigen Gewinnmargen* (FAZ 1995, o.S., IDS-Korpora). Vgl. auch den Kinderbuchtitel Stein-

substantivischen Zusammensetzungen ein Nebeneinander von Wortformen und einer einzigen unparadigmischen Form des Typs *Kindheits*. Diese eine Form hat das auch paradigmisch vorkommende Element *s*.¹⁰

Betrachten wir nun die Bedeutung solcher Zusammensetzungen.

Flexionselemente markieren Bedeutung, unter anderem Zugehörigkeit (*Freundin des Königs*), bei Sachen speziell auch Besitz (*Mantel des Königs*)¹¹ sowie Vielzahl (*Regal für Bücher*). Davon werden Fugenelemente als bedeutungsleer abgegrenzt. Aber welche Elemente sind bedeutungsleer?

- **Position 1** sieht Elemente wie *es* oder *en* generell als bedeutungsleer, denn „die Funktion der Flexionsendungen übernehmen die Fugenelemente häufig gerade nicht“ (Fuhrhop 1998, S. 191).¹² Wie erkennt Position 1 das? Ortner et al. (1991, S. 98) erkennen es so: „Dass *-er-* in der Kompositionsfuge nicht einfach als Pluralzeichen interpretiert werden kann, bezeugen schon Zusammensetzungen wie *Hühnerkeule*, *Kinderjacke* und *Rinderbrust*. Das Bestimmungswort bezieht sich bei ihnen auf ein einzelnes Huhn, Kind, Rind.“ Nebenbei bemerkt: Wie würden Ortner et al. das für *Hühnerkeulen*, *Kinderjacken* oder *Rinderbrüste* ausdrücken? Und was würde daraus folgen?
- **Position 2** meint, dass ein Erstglied nicht „ohne Wert ist, sondern auch stilistische, pragmatische und textuelle Funktionen erfüllt“ (Kauffer 1993, S. 420). So interpretiert Position 2 *Anwaltskammer* als ‘Kammer des Anwalts’ und versteht *Anwalt* als Gattungsname, nicht als singuläre Bezeich-

höfel, Andreas (2008): Rico, Oskar und die Tieferschatten. Hamburg. *Tieferschatten* bezeichnet besonders dunkle Schatten.

¹⁰ Mitunter wird außerdem *en* etwa in *Dokumentenmappe* oder *Zitatenschatz* als unparadigmisch angegeben, so bei Ortner et al. (1991, S. 77); es gehört aber klar zum Paradigma dieser Wörter: *eine Mappe mit Dokumenten*. Andere Formen, die heute im Paradigma nicht vorkommen, wie in *Sternenlicht*, *Mondenschein* oder *Schelmenstreich*, lassen sich durch historische Genitivformen erklären. Vgl. unter vielen anderen Paul (1975, S. 336) und Ramers (1997, S. 36). Dies wird oft übersehen, „because most research on linking elements is synchronically oriented“ (Nübling/Szczepaniak 2008, S. 6).

¹¹ Zifonun (2010) entwickelt eine funktional begründete „Alternative zu der unbefriedigenden traditionellen Klassifikation des deutschen Genitivattributs“ (ebd., S. 149).

¹² So auch Eschenlohr (1999, S. 206) und Wegener (2003, S. 446).

nung einer einzelnen Person.¹³ *Kinderjacke* interpretiert Position 2 als ‘Jacke für Kinder’. Es gibt nämlich immer Möglichkeiten, die Form des Erstglieds in eine Interpretation zu integrieren. Sogar Verfechter der Position 1 wie Ortner et al. (1991, S. 110) sehen in Bildungen wie *Gästebuch* einen Plural, weil „der Bezug auf eine Vielheit verdeutlicht werden soll“. Eisenberg (2004, S. 239) räumt bei *Schriftenverzeichnis* und *Zeitenfolge* „doch eine Tendenz zur Wahl der Fuge nach der Pluralbedeutung“ ein. Auch Fuhrhop (1998, S. 193) findet bei *Bücherregal*, *Städtetag* und *Mäusegift* „die Interpretation der Pluralformen als Mehrzahl durchaus angemessen“. Nübling/Szczepaniak (2008, S. 3) finden, dass *Blumen* in „*Blume + n + strauß* ‘bouquet of flowers’, however, can be analysed as a plural“. Wo Wegener (2005, S. 175) *Hundehütte* als eine Zusammensetzung sieht, die „keine Pluralbedeutung aufweist“, und Eisenberg (2004, S. 236) von einem „semantisch falschen Plural“ etwa bei *Gänsebraten* spricht, entdeckt Varsami (2007, S. 9): „Eine Hundehütte ist einfach eine Hütte für Hunde (wenige würden wohl sagen ‘Eine Hütte für den Hund’). Der Plural in *Schweinebraten* scheint auf den ersten Blick nicht sinnvoll gebildet. Doch auch hier ist eine Erklärung denkbar einfach: Wenn man aus Schweinen (Plural!) Wurst machen kann, dann lässt sich aus Schweinen auch ein Braten braten. Schneide ich einen einzelnen Pilz in eine Omelette, sage ich trotzdem, dass Pilze drin sind.“¹⁴ Es gibt also Interpretationsspielraum. Wie großzügig möchte man darin spielen? Heringer (2009b, S. 95) entscheidet so: „Alle Flexive, die deutbar sind, zählen wir nicht zu den Fugenzeichen“. Position 2 hält Deutbarkeit für ein weites Spielfeld, in dem alle paradigmatischen Elemente deutbare Flexionselemente sind.

Bedeutung schillert außerdem in die Diskussion um die Unregelmäßigkeiten der Formen hinein: So verwenden wir unregelmäßig kein flektiertes Erstglied, wo das aus semantischen Gründen nahe läge: Eine Bücherkiste bietet Platz für Bücher, aber dort, wo noch vielzähliger Bücher verkauft werden, befinden wir uns in einer Buchhandlung.

¹³ So meint auch das werbewirksame *Advocard ist Anwalts Liebling* bekanntlich nicht den Liebling eines bestimmten einzelnen Anwalts und im folgenden Beleg geht es ebenfalls um den Seemann an sich: *Des Seemanns Qual ist digital – Früher brauchte ein Skipper noch solides Handwerk und einen Sextanten, um einen Ozean zu überqueren. Seitdem es GPS gibt, kann jede Landratte segeln* (<http://sz-magazin.sueddeutsche.de>, September 2009).

¹⁴ Und warum ist der Singular in *Freundeskreis* eigentlich „semantisch falsch“ (Eisenberg 2004, S. 237)? Es ist der Kreis eines Freundes. Es ist der Kreis, den ein Freund um sich schart.

- **Position 1** erwartet eine Regelmäßigkeit, wie sie für Flexionselemente obligatorisch ist, und leitet aus der Unregelmäßigkeit ab, dass uns flektierte Erstglieder verdunkelt sind. Denn würden wir uns der Flexionsform bewusst sein, müsste zum Beispiel „in einigen Fällen die Abwesenheit des Genitiv-s- begründet werden: wieso heißt es *Apfelkern* und nicht *Apfelskern*?“ (Fuhrhop 1998, S. 193). Auch wer *Hundeleine* oder *Mäusezahn* als Zusammensetzungen mit *e*-Fuge versteht, wundert sich, dass solche Fugen „keineswegs bei allen Substantiven dieser Art stehen“, zum Beispiel nicht bei *Fuchsschwanz*, *Wolfsrachen* und *Schafstall* (Eisenberg 2004, S. 238f.), denn es heißt üblicherweise nicht *Füchseschwanz*, *Wölferachen* und *Schafestall*.
- **Position 2** sieht jedes System als ein System offener Optionen. Die Option, Wörter mit verschiedenen Wortformen zu bilden, nutzen wir mal so, mal so. Dass wir mal so, mal so verfahren, ist für Position 2 kein Argument für oder gegen irgendetwas.

3. (In)Fragestellung

Es ist nicht unbedingt nötig, vielleicht auch gar nicht möglich zu entscheiden, ob Position 1 oder Position 2 weiter führt. Ich muss aber entscheiden, wie ich in dieser Pilotstudie über mein Thema spreche. Für diese Studie habe ich mich für Position 2 entschieden. Danach ist das Erstglied in substantivischen Zusammensetzungen entweder eine paradigmatische Wortform (*Tageslicht*) oder eine Form mit unparadigmatischem *s* (*Kindheitstraum*). Nur Letzteres nenne ich Fugenelement. Ausschließlich Fugenelemente wie in *Kindheitstraum* erfüllen alle traditionellen Definitionsmerkmale von Fugenelementen, besonders das Merkmal der Fugenfüllung, des Fugenkitts, und das Merkmal der Bedeutungsleere.

Ich schließe mich damit unter anderem der Definition von Darski (1996, S. 53) an, nach dem als Fugenelement ein Element verstanden wird, das „vor der letzten Wortform in der Zusammensetzung steht und nicht zu der davor stehenden Wortform [...] gerechnet werden kann. Fugenelemente sind zum Beispiel: [...] *s* in *Geschichtsbuch* und *Wohnungsamt*. Keine Fugenelemente sind nach dieser Definition [...]: *er* in *Kinderbett*; (*e*)*s* in *Tageslicht*, *Himmelskörper*; *en* in *Dornenkrone*; *ens* in [...] *Herzensfreude*; *e* in *Tagebuch* usw., weil sie als Bestandteile der Bestimmungswörter zu betrachten sind“.

Vielleicht hat diese Sichtweise auch für den Laien Vorteile. Vielleicht hat er das Gefühl der besseren Übersicht und muss nicht verwirrt anfragen wie dieser nette, durchaus typische Anfrager.

Sehr geehrte Damen und Herren,

ich bitte Sie um Auskunft in folgender Frage:

Das „Gesetz- und Verordnungsblatt der Evangelischen Landeskirche in Baden“ gebraucht bei Ausschreibungen immer wieder Terminus und Schreibweise „Interessensmeldungen“ oder „Interessensbekundungen“. Meinem Sprachempfinden nach ist das Fugen-s in diesem Fall störend, weil (Singular von Interesse vorausgesetzt) bereits ein Fugen-n verwendet wird. Eigentlich könnte es ja sogar „Interessemeldung“ heißen, oder? Ich wäre Ihnen dankbar, wenn Sie mir kurz Auskunft geben könnten, damit ich sie gegebenenfalls an den Redakteur des Blattes weiterreichen kann.

Mit freundlichen Grüßen

D.Z., Pfarrer i.R.

Position 1 sieht hier eine Vielzahl irgendwie verboten addierter Fugenelemente. Position 2 sieht eine reguläre paradigmatische Pluralform *Interessen* und ein reguläres unparadigmatisches Füllungsfugen-*s*. Auch löst Position 1 nicht besser als Position 2, dass wir alle gerne nach Regeln lernen würden. Beide Positionen gehen ja von einem gewissen unerklärlichen Chaos aus; bei beiden Positionen bleibt das Problem *Tagtraum, Tageslicht, Tagedieb*.

„Betrachtet man nämlich diesen Datenbereich, so scheint auf den ersten Blick Goethes Ausspruch über die Grammatik aus ‘Dichtung und Wahrheit’ recht passend“ (Ramers 1997, S. 33). Goethe sprach aus: „Die Grammatik mißfiel mir, weil ich sie nur als ein willkürliches Gesetz ansah; die Regeln schienen mir lächerlich, weil sie durch so viele Ausnahmen aufgehoben wurden, die ich alle wieder besonders lernen sollte.“ Auch Jean Paul (1819, S. 14) spricht so: „Dem Anschein nach scheint nichts regelloser als die Art, auf welche unsere Sprache in den Doppelwörtern das Bestimmungswort mit dem Grundwort verknüpft.“ Auf seinen zweiten Blick aber scheint Ramers (1997, S. 33) die Formulierung von Regeln für die Verwendung von Wortformen und Fugenelementen „ein zwar schwieriges, aber keineswegs lächerliches Unterfangen“. Nützt dabei eine große, intelligent genutzte Datenbasis mit Blick auf Varianten?

4. Korpus und Analysemethode

*Beendet den langen Marsch
durch die Intuitionen!
Korpora intelligent nutzen!*
(Heringer 2009a, S. 31)

Eichinger (2005, S. 369) konstatiert, dass „Fragen der empirischen Basis wichtiger geworden sind“. „Das Zutrauen zur Dokumentation des Sprachgebrauchs als zu befragende Instanz hat zweifellos zugenommen, seit uns hier elektronische Korpora eine gänzlich neue Art der Übersicht erlauben – und damit auch eine andere Rahmung für Experimente.“ (Eichinger 2009, S. 5).

In dieser Studie will ich experimentieren. Ich will vor allem herausfinden, ob ich variierende Wortformen und Fugenelemente in Zusammensetzungen mit Hilfe großer Textkorpora sprachrealitätsnäher beschreiben kann. Ich befrage zunächst bisherige Untersuchungen. Wie werden Datenbasen dort erstellt und genutzt?

Die Hauptdokumentation der deutschen Substantivzusammensetzung ist Ortner et al. (1991). Nach wie vor gibt es keine vergleichbar breit angelegte Dokumentation. Bei ihrem Korpus handelt es sich „einerseits um Texte sehr verschiedener Textarten, besonders aus der Presse und Literatur (Lyrik und Prosa) und aus vielfältigen Sachdarstellungen, andererseits um das Material aus Wörterbüchern, vor allem aus dem Wörterbuch der deutschen Gegenwartssprache (von Klappenbach/ Steinitz)“ (Ortner et al. 1991, S. XXXIII).

Im Rahmen meiner Studie kann ich viele Fragen nicht beantworten, aber fragen möchte ich trotzdem. Zum Beispiel: Welches Wörterbuch ist optimal zum Aufbau eines Korpus? Das dickste? Das durchstrukturierteste? Das mit den meisten authentischen Quellen? Das aus einer Hand? Vielleicht, wenn es ein Spezialwörterbuch zur Varianz von Wortformen und Fugen in deutschen Zusammensetzungen gäbe? Oder verfälscht die Einbeziehung von Wörterbüchern grundsätzlich? Wörterbucheinträge sind ja nicht produzierte Sprache, sondern Auswahl, Auswertung und nutzungsspezifische Anordnung vorgefundener Sprache. Ist nicht ein Wortbildungsforscher, der Wortbildung aus Wörterbüchern lesen will, wie ein Schmetterlingsforscher, der alles über Schmetterlinge aus aufgespießten Glaskastenexemplaren lesen will? Zum Beispiel werden Erstgliedformen in eine wörterbuchtaugliche Form gebracht; bestenfalls ist das die häufigste, schlimmerenfalls die dem Wörterbuchautor

bekannteste. Und dann beruhen gerade frühere Wörterbücher auf Verzettlung und wie ich aus meiner Praxis als Verzettlerin weiß: Verzettelt werden ungewöhnliche Ereignisse, nicht das Gros der Normalereignisse.¹⁵ Verzerrt das das Sprachbild, das uns Wörterbücher zeigen? Oder spiegeln Wörterbücher die selbe Sprache wie andere produzierte Sprache auch? Kann man sie lesen als produzierte Sprache von Wörterbuchautoren? Aber warum dann ausgerechnet Wörterbuchautoren?

Kürschner hat deutsche und dänische Zusammensetzungen aus Substantiven „nach dem Zufallsprinzip aus Internetpublikationen deutscher und dänischer Tageszeitungen extrahiert“ (Kürschner 2003, S. 58). Ist das Zufallsprinzip brauchbarer als die verzettelte Auswahl bei Ortner et al. (1991)? Nützt es, wenn wir uns als Linguisten unserer verzettelnden Hand bewusster sind? Sehen wir überhaupt „unsere eingeschränkte Fähigkeit, Reihen von Fakten zu betrachten, ohne eine Erklärung in sie hineinzuwoben oder, was dasselbe bedeutet, gewaltsam logische Verknüpfungen, einen Beziehungspfeil, zwischen ihnen herzustellen“ (Taleb 2010, S. 89)? Befreit uns ein Rechner von unserer Hand? Ist er uns überlegen?

Breit gefächerte Korpora gelten allenthalben als optimal. Dem Quellenverzeichnis nach zu urteilen, ist das Korpus von Ortner et al. breit gefächert: Verzettelt wurden Romane (unter anderen von Böll, Goethe, Kästner, Lovecraft, Mann und Zweig), Gedichte (unter anderen von Benn, Celan, Heym und Höllerer), Zeitschriften (wie *Hör zu*, *Spiegel* und *musik express*), Fachzeitschriften (wie *Sprachpflege*, *Germanistik* und *Theater heute*) und populäre Fachliteratur (mit Titeln wie *Grundlagen der organischen Chemie*, *ABC Biologie*, *Schnellkochbuch*).

Brauchen wir eine so breite Fächerung? Brauchen wir eine noch breitere Fächerung? Brauchen wir Amtstexte, Juristisches, Bedienungsanleitungen, private Briefe, Tagebücher, lässige Internetblogs? Brauchen wir 10% Hochliteratur und 11% Internetblogs und 15% Kinderbücher? Wonach sollten sich prozentuale Anteile richten? Nach dem, was wir im Durchschnitt so sagen und schreiben? Nach dem, was wir täglich im Durchschnitt hören und lesen? Nach der Häufigkeit der Texte? Nach Auflagenzahl? Nach der Wichtigkeit der Texte für unsere Kulturgemeinschaft?

¹⁵ „Wenn wir versuchen, uns Informationen über die Welt um uns herum zu besorgen, neigen wir dazu, uns von unserer Biologie leiten zu lassen, und unsere Aufmerksamkeit richtet sich wie von selbst auf das Sensationelle – nicht so sehr auf das Relevante, sondern auf das Sensationelle.“ (Taleb 2010, S. 117).

Ich recherchiere Sprachliches gerne in Google,¹⁶ weil die Googlewelt groß und bunt ist, aber auch Google ist natürlich nicht die ganze Sprachwelt. Brauchen wir die ganze Sprachwelt? Brauchen wir zumindest repräsentative Ausschnitte aus dieser Sprachwelt? Kann man überhaupt die Sprachwelt einfangen oder doch wenigstens abbilden?

Bilden breit gefächerte Texte Sprache besser ab als Texte aus einem speziellen Genre? Kauffer (1993) hat ein themenspezifisches Korpus aus Speisekarten mit über 2 000 substantivischen Zusammensetzungen aufgebaut. Reicht das, um Grundfragen zur Verwendung von Wortformen und Fugenelementen in Zusammensetzungen zu beantworten?

Zumindest ist erkennbar, dass die Trefferquote schon bei kleinen Teilkorpora relativ hoch ist. Die von Noah Bubenhofer für meine Studie zusammengestellte DeReKo-Schnittstellenanalysen berücksichtigen zum Beispiel ein Lufthansa-Bordbuch, das zum Beispiel über 1 000 Schnittstellen mit *s* liefert. Führen verschiedene und verschieden große Teilkorpora zu verschiedenen Ergebnissen? Oder ist es egal, wie groß man einen Sprachausschnitt wählt.¹⁷

Geprüft habe ich das in Tabelle 1 an Erstgliedern auf *ung* plus *s* (*Änderungsschneiderei*, *Ahndungsvermögen*). Welchen Anteil haben sie an allen Erstgliedern mit *s*? Recherchiert habe ich in unterschiedlichen Teilkorpora unterschiedlicher Größe. Ich gebe sie hier in aufsteigender Größe wieder. Gerechnet wurden Tokens.

	Hannoversche Allgemeine	Lufthansa Bordbuch	Thomas Mann	Reden Bundes- tag	Computer Zeitung	Die Zeit
Erstglieder + <i>s</i>	219	1 074	12 303	20 580	90 428	501 756
davon Erstglieder auf <i>ung</i> + <i>s</i>	26,0%	26,6%	16,0%	35,4%	31,8%	6,6%

Tabelle 1

¹⁶ Zu modifizierteren Sichten auf Google als linguistisches Korpus vgl. etwa Kilgariff (2006).

¹⁷ „Wir können also viel aus Daten erfahren – allerdings nicht so viel, wie wir erwarten. Manchmal kann eine große Datenmenge bedeutungslos sein, in anderen Fällen eine einzige Information von sehr großer Bedeutung. 1 000 Tage können nicht beweisen, dass wir recht haben, doch ein einziger Tag kann beweisen, dass wir uns irren.“ (Taleb 2010, S. 81).

Je nach Ausschnitt schwankt der Anteil der Erstglieder auf *ung* plus *s* also. Die meiner Studie zugrunde liegenden DeReKo-Schnittstellenanalysen zeigen insgesamt ein Verhältnis von rund 30 000 000 zu 7 000 000. Das macht ein knappes Viertel aus (23%), ähnlich wie in der Hannoverschen Allgemeinen (26%), aber ganz anders als in Die Zeit (6,6%) oder bei Thomas Mann (16,0%). Zufall?¹⁸ Und ist wenigstens die 30-Millionen-Menge aussagekräftig genug, um etwas über die Sprache zu sagen? Wie viel brauchen wir, um allgemeine Aussagen machen zu können? Welcher Ausschnitt sagt etwas aus über den Anteil allgemein? Und not least: Was ist das überhaupt, über das wir allgemeine Aussagen machen?

Krott et al. (2007) arbeiten mit Celex. Laut dessen Betreibern Baayen, Piepenbrock und Gulikers „for German, the total number of lemmas included is now 51,728“ (<http://www ldc.upenn.edu/Catalog/topten.jsp>, April 2010). Das Korpus von Ortner et al. „umfasst 62456 verschiedene Substantivkomposita“ (Ortner et al. 1991, S. XXXIII). Wäre noch mehr besser? Wenigstens ist ein Mehr heutzutage leicht zu bewerkstelligen. Vermutlich machen wir mit einem Mehr nichts falsch. Solange wir nicht falsch interpretieren ...

Meiner Studie liegen die von meinem IDS-Kollegen Noah Bubenhofer erarbeiteten DeReKo-Schnittstellenanalysen mit rund 177 000 000 Analysen zugrunde. Worauf sie basieren und wie genau sie strukturiert sind, beschreibt er selbst im folgenden Exkurs.

¹⁸ Das Type-Token-Thema wird in Kapitel 5 gesondert behandelt. Hier kann ich aber schon mal zur Verwirrung fußnotieren, dass Rechnungen mit Types ganz andere Prozente ergeben: So liegt der Anteil der *ung*-Erstglieder mit *s* an allen Erstgliedern mit *s* in der Hannoverschen Allgemeinen bei 32,0%, im Lufthansa Bordbuch bei 36,5% und im Thomas-Mann-Korpus bei 35,9%.

Zur automatischen Analyse von Zusammensetzungen

Exkurs von Noah Bubenhofer

Grundlage der Schnittstellenanalysen bildet eine Datenbank von Zusammensetzungen im DeReKo,¹⁹ die nach den darin vorkommenden Schnittstellen-Typen klassifiziert sind. Die Daten wurden mittels einer automatischen Analyse der morphosyntaktisch annotierten Version des DeReKo erhoben, die in diesem Exkurs beschrieben wird.

Grundlage: Die annotierte Version des DeReKo

Das DeReKo steht in verschiedenen morphosyntaktisch annotierten Versionen zur Verfügung.²⁰ Zum Aufbau der Schnittstellen-Datenbank wurde die mit dem „MachineSe“-Tagger der Firma „Connexor“²¹ annotierte Version verwendet, denn der Tagger nimmt bei Zusammensetzungen eine morphologische Analyse vor, wie im folgenden Ausschnitt der Annotation von *Haushaltsgeräte* zu sehen ist:

```
<token pos="3671970" len="15">
  <text>Haushaltsgeräte</text>
  <lemma>haushalt gerät</lemma>
  <tags syntax="@NH" morpho="N" sub2="PL"/>
</token>
```

Das Token *Haushaltsgeräte* wird also im Tag `<lemma>` in die Glieder *haushalt* und *gerät* aufgetrennt. Die Schnittstelle *s* wird dabei ignoriert.

Automatische Extraktion der Zusammensetzungen

Mit einem speziell für diesen Zweck programmierten Perl-Skript²² werden die Connexor-getaggten Korpusdaten ausgelesen. Berücksichtigt werden alle als Substantiv klassifizierten Tokens, die gemäß Lemma-Angabe in der Annotation aus mehreren Gliedern bestehen. Die aufgeführten Glieder (im Beispiel

¹⁹ Deutsches Referenzkorpus des Instituts für Deutsche Sprache (<http://www.ids-mannheim.de/kl/projekte/korpora/>)

²⁰ Vgl. <http://www.ids-mannheim.de/kl/projekte/korpora/annotationen.html>

²¹ Vgl. <http://www.connexor.com>

²² *Perl* ist eine Programmiersprache, die sich besonders für die Verarbeitung von Textdokumenten eignet. Vgl. Wall et. al. (2001).

oben: *haushalt* und *gerät*) werden nun mit dem tatsächlichen Token verglichen (*Haushaltsgerät*). Es wird berechnet, welche Operationen notwendig sind, um zum Token zu gelangen. Dabei werden die folgenden Fälle in der genannten Reihenfolge unterschieden:

1. Die Glieder lassen sich über eine der vordefinierten Schnittstellen *s*, *en*, *e*, *n*, *er*, *es*, *a* verbinden (*Pferdewagen*).
2. Die Glieder lassen sich ohne jegliches weitere Element miteinander verbinden (*Abfallwärme*).
3. Eine beliebige andere Zeichenkette ist notwendig, um die Glieder zu verbinden (*Herzengüte*).
4. Eine Verkürzung des ersten Glieds in Verbindung mit einem der vordefinierten Schnittstellen ist notwendig (*Hilfskonstruktion* [*Hilfe* → *Hilf* + *s*]).
5. Verkürzung des ersten Substantivs ohne weitere Veränderungen (*Grenzdienst*).
6. Verkürzung des zweiten Substantivs, allenfalls inklusive Hinzufügung eines Flexionselements. Diese Variante ist für den Zweck der Analyse eigentlich irrelevant und wird deshalb bei der Zählung der Kategorien nicht differenziert.
7. Wenn keine der genannten Fälle zutrifft, wird die Zusammensetzung als unanalysierbar markiert, wobei eine Reihe von häufigen Spezialfällen berücksichtigt werden, die auf orthographische Unterschiede zwischen Lemma und Token wie *-graph-* vs. *-graf-*, *-photo-* vs. *-foto-* etc. zurückgehen.

Generell ignoriert werden Pluralumlaute wie in *Ärtekongress* (*arzt* + *kongress*): Vor der Analyse werden Umlaute generell ersetzt und die Zusammensetzung *Ärtekongress* würde dann als Fall *Arzt* + *e* + *Kongress* klassifiziert.

Bei Zusammensetzungen, die aus mehr als zwei Gliedern bestehen, wird jede Schnittstelle separat klassifiziert und als Paar von jeweils einem Erst- und einem Zweitglied gespeichert. Dabei wird keine Hierarchisierung im Prozess der morphologischen Analyse vorgenommen.

Neben der Klassifizierung der Zusammensetzungen wird zu jedem gefundenen Token auch die genaue Fundstelle (Korpus und Byteposition im Korpus) abgelegt.

Erkennung von paradigmatischen s-Schnittstellen

Für die Auswertung interessierte die Differenzierung zwischen (e)s-Schnittstellen, die Teil des Flexionsparadigmas (Genitiv-Form) sind und den restlichen. Um diese Differenzierung maschinell vorzunehmen, wurde überprüft, ob das Erstglied im Genitiv eine *es*- oder *s*-Endung aufweist. Als Quelle der Flexionsparadigmata wurde das deutschsprachige „Wiktionary“²³ ausgelesen, die Wörterbuchvariante der Wikipedia. Im Wiktionary ist zu jedem Substantiv eine Flexionstabelle der folgenden Art verfügbar:²⁴

Kasus	Singular	Plural
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus dem Hause	den Häusern
Akkusativ	das Haus	die Häuser

Tabelle I

Diese Flexionstabellen sind formal regelhaft codiert und können deshalb aus dem Datenbankdump²⁵ mittels folgender UNIX-Befehle und Perl-Skripte ausgelesen und es kann damit eine Liste aller Substantive mit *-es/-s*-Endung im Genitiv erstellt werden:

```
egrep "\|Wessen\? \ (Einzahl\)=.+(\ (e\ )s|es|s)$" dewiktionary-
20100210-pages-articles.xml > SubstantiveGenitivS.txt
perl -pi -e 's/\|Wessen\? \ (Einzahl\)=(des|\
(des\ )|der)?\s?(.+)/\2/' SubstantiveGenitivS.txt
perl -pi -e 's/<\|/?.+?>/ /g' SubstantiveGenitivS.txt
perl -pi -e 's/\s+/\n/g' SubstantiveGenitivS.txt
perl -pi -e 's/(.+?) (\(|\(|e(|\)|))s/\ls\n\les/g' Substantive-
GenitivS.txt
perl -pi -e 's/\((? (der|des)\ )?\n//g' SubstantiveGenitivS.txt
sort SubstantiveGenitivS.txt | uniq > SubstantiveGenitivS.
sorted.txt
```

²³ Vgl. <http://de.wiktionary.org/>

²⁴ Vgl. <http://de.wiktionary.org/wiki/Haus>

²⁵ Erhältlich über <http://dumps.wikimedia.org/backup-index.html>

Im Folgenden werden alle Zusammensetzungen, deren Erstglied in der Liste erscheinen, mit der Information „paradigmisches s“ kodiert.

Erkennung von Substantiv-Substantiv-Zusammensetzungen

Bei der Auswahl der Zusammensetzungen aus dem Korpus wurden alle Zusammensetzungen berücksichtigt, die vom Tagger als Substantiv klassifiziert wurden. Allerdings handelt es sich dabei nicht zwangsläufig um Substantiv-Substantiv-Zusammensetzungen (*Kinderarzt*), sondern auch um solche, bei denen das Erstglied einer anderen Wortart angehört (*Abwärtstrend*). Um zu überprüfen, ob es sich beim Erstglied um ein Substantiv handelt, wurden die mit der Xerox FST Linguistic Suite annotierten Daten des DeReKo verwendet. Der Vorteil des Xerox-Taggers liegt darin, dass bei Zusammensetzungen auch die Glieder mit Wortarten annotiert sind.²⁶ Ein Perl-Skript extrahiert alle Substantive aus dem Xerox-annotierten Korpus. Diese Substantivliste wird bei der Extraktion verwendet, um bei jeder Zusammensetzung zu überprüfen, ob dessen Bestandteile in der Substantivliste vorkommen.

Datenbank

Die in den oben beschriebenen Schritten erzeugten Informationen werden in einer Datenbank abgelegt, wobei zwei unterschiedliche Tabellen verwendet werden:

1. Tabelle der Schnittstellen auf Token-Basis: Pro gefundener Zusammensetzung bestehend aus n Gliedern werden $n-1$ Einträge der Erst-/Zweitgliedkombinationen mit den Angaben zum Schnittstellentyp, Fundort etc. erstellt. Diese Tabelle enthält 176 405 282 Schnittstellen.
2. Tabelle der Schnittstellen auf Type-Basis: Erst-/Zweitgliedkombinationen mit gleichem Schnittstellentyp werden zusammengefasst und es wird zusätzlich vermerkt, wie oft diese Kombination in den Korpora vorkommt. Diese Tabelle enthält 8 699 856 Erst-/Zweitgliedkombinationen mit jeweils gleichem Schnittstellentyp.

²⁶ Dass für die maschinelle Extraktion von Zusammensetzungen trotzdem auf den Connexor Machinese-Tagger zurückgegriffen wurde, hat verschiedene Gründe. Der wichtigste Grund liegt darin, dass der Xerox-Tagger zu jeder Zusammensetzung alle möglichen morphologischen Analysen liefert, ohne eine Disambiguierung vorzunehmen. Der Connexor-Tagger gibt nur eine Lösung aus, die die wahrscheinlichste Lösung zu sein scheint. Leider klärt die Dokumentation aber nicht auf, ob diese Vermutung korrekt ist.

Um die Datenbank benutzerfreundlich abfragen zu können, wurde darüber hinaus eine webbasierte Benutzeroberfläche erstellt:

DeReKo Schnittstellenanalysen-Datenbank

Version 1.14, vgl. die [Release Notes](#)

Feld	Wert	Aggregiert
Kompositum:	<input type="text"/>	<input type="checkbox"/>
Erstglied:	<input type="text" value="fuss"/>	<input type="checkbox"/>
Zweitglied:	<input type="text" value="ball"/>	<input type="checkbox"/>
Schnittstelle*:	Bitte wählen... <input type="checkbox"/> Verkürzungen miteinbeziehen	<input type="checkbox"/>
Frequenz:	> <input type="text"/>	
Stichprobe:	100% der Daten <input type="button" value="v"/>	
Erstglied Wortart:	unbestimmt <input type="button" value="v"/>	
Erstglied: Genitiv-S im Paradigma:	egal <input type="button" value="v"/>	
Sortierung nach:	Erstglied <input type="button" value="v"/> <input checked="" type="radio"/> aufsteigend <input type="radio"/> absteigend	
Maximale Rückgabemenge:	10.000 Datensätze <input type="button" value="v"/>	

* Achtung: Die Schnittstellenanalysen-Datenbank ist automatisch generiert. Was hier als "Schnittstelle" ("Fuge") bezeichnet wird, muss deshalb nicht zwingend auch linguistisch eine Schnittstelle/Fuge sein, da der Algorithmus auch fehlerhafte Analysen machte oder bereits die Daten fehlerhaft morphologisch analysiert waren. Vgl. dazu auch die [Dokumentation](#).

[Hilfe](#)

Korpusauswahl

Achtung: Abfragen über alle Korpora sind am schnellsten. Abfragen über wenige Korpora sind ebenfalls relativ schnell. Aber Abfragen über viele (aber nicht alle) Korpora können sehr lange dauern!

☒ Alle Korpora
☐ Aggregieren nach Korpus

Automatische Auswahl:

Zeitungungen <input type="checkbox"/> a (St. Galler Tagblatt) <input type="checkbox"/> h (Berliner Zeitung)	Andere Korpora <input type="checkbox"/> bih (Hrsg.-Texte zu Victor Klemperer) <input type="checkbox"/> hi1z8 <input type="checkbox"/> hi1z9 <input type="checkbox"/> hi1zf
--	---

Abb. I: Benutzeroberfläche der DeReKo-Schnittstellenanalysen-Datenbank

Die Datenbank erlaubt folgende Abfragen, die kombinierbar sind:

1. Suche nach bestimmten Zusammensetzungen unter Verwendung von Platzhaltern.
2. Suche nach Zusammensetzungen, die ein bestimmtes Erst- oder Zweitglied aufweisen; es können ebenfalls Platzhalter verwendet werden.
3. Suche nach Zusammensetzungen oder Erst-/Zweitgliedern mit spezifischen Schnittstellen.
4. Zusätzlich Angabe einer Mindestfrequenz, mit der die Zusammensetzung in der Korpusgrundlage verwendet wird.
5. Auswahl einer Zufallsstichprobe aus allen Datensätzen.

6. Einschränkung auf Erstglieder, die als Substantiv annotiert sind.
7. Einschränkung der Erstglieder auf solche mit paradigmatischen *-s/-es*-Endungen.
8. Sortierung der Datensätze alphabetisch, nach Frequenz etc.
9. Wahl der Datengrundlage: Auswahl einer Anzahl bestimmter Teilkorpora des DeReKo.
10. Aggregationsfunktionen: Die zurückgegebenen Zusammensetzungen bzw. Erst-/Zweitglieder oder Kombinationen davon können nach den Kriterien Zusammensetzung, Erstglied, Zweitglied und Schnittstelle gruppiert (aggregiert) werden.

Beispiel: Bei der Suche nach dem Erstglied *kind*, aggregiert nach Schnittstelle wird eine Liste der Schnittstellen in Verbindung mit *kind* als Erstglied mit den jeweiligen Frequenzen ausgegeben.

Mit der Trefferliste werden jeweils zwei Frequenzwerte angegeben: Die Anzahl der Types („Treffer“) und der Tokens („Gesamtfrequenz der Treffer“). Beispiel: Die Suche nach dem Lemma *kind* als Erstglied, aggregiert nach Zweitgliedern, erzeugt eine Liste aller unterschiedlichen Zweitglieder, die zusammen mit dem Erstglied *kind* vorkommen. Die Anzahl Treffer beträgt 3 622, es gibt also 3 622 unterschiedliche Zweitglieder. Die Gesamtfrequenz dieser Treffer beträgt aber 851 438, d.h. die 3 622 unterschiedlichen Zweitglieder kommen im Korpus insgesamt 851 438 mal vor.

Probleme

Es gibt eine ganze Reihe von Fehlerquellen, die zu Problemen in der Schnittstellen-Datenbank führen. Grundsätzlich können bereits Fehler bei der Tokenisierung (Erkennung der Wortgrenzen), der Korpuszusammenstellung (z.B. Textdubletten) und anderer Verfahren der Korpusaufbereitung einen Einfluss haben. Auf solche Probleme soll an dieser Stelle nicht eingegangen werden, da sie auch die Suche im DeReKo betreffen und damit genereller Natur sind und nicht so einfach behoben werden können.

Im Folgenden soll auf ein paar Beobachtungen aufmerksam gemacht werden, die bei der Erstellung der Schnittstellen-Datenbank gemacht worden sind.

Fehlerhafte morphologische Analyse des Connexor-Taggers

Wie jeder Wortarten-Tagger weist auch der Connexor-Tagger Fehler auf. Fehler im Bereich der morphologischen Analyse wirken sich auf die automatische Bestimmung von Schnittstellen aus. Da es sich beim Connexor-Tagger um ein kommerzielles Produkt handelt, ist die Funktionsweise intransparent und Korrekturen sind schwierig.

Inkonsistenz bei der Analyse von Zusammensetzungen:

Embryo(nen)forschung

Problem:

Embryonenforschung und *Embryoforschung* wird je und in sich unterschiedlich analysiert. So wird *Embryoforschung* korrekt als Zusammensetzung analysiert:

```
<token pos="81043469" len="15">
<text>Embryoforschung</text>
<lemma>embryo forschung</lemma>
<tags syntax="@NH" morpho="N"/>
</token>
```

Hingegen wird *Embryonenforschung* nicht als Zusammensetzung erkannt. Zudem scheint die Lemmatisierungskomponente manchmal das Token als (komplexes) Lemma zu erkennen, manchmal nicht, was sich in der Groß-/Kleinschreibung innerhalb des Tags erkennen lässt:

```
<token pos="632878274" len="18">
<text>Embryonenforschung</text>
<lemma>Embryonenforschung</lemma>
<tags syntax="@NH" morpho="N" sub1="Prop"/>
</token>
--
<token pos="777958976" len="18">
<text>Embryonenforschung</text>
<lemma>embryonenforschung</lemma>
<tags syntax="@NH" morpho="N"/>
</token>
```

Schreibvarianten mit Bindestrich (*Embryo(nen)-Forschung*) werden hingegen korrekt analysiert.

Auswirkung auf die Analyse:

In der Datenbank ist *Embryonenforschung* nicht auffindbar, da es beim Auslesen der getaggtten Daten nicht als Zusammensetzung erkannt wurde. Die Wortform ist jedoch in DeReKo sehr frequent (1 300 Vorkommen mit flektierten Varianten). Die Schnittstelle *nen* zu *embryo* ist also viel häufiger, als die Datenbank ausgibt.

Es ist momentan nicht abschätzbar, wie viele ähnliche Fälle von (teilweise) falsch analysierten Zusammensetzungen es gibt. Es ist jedoch anzunehmen, dass allgemein eher seltene Lexeme falsch analysiert sind.

Inkonsistenz bei der Analyse von Zusammensetzungen: *Wille(n)*-

Problem:

Zusammensetzungen, die mit *Wille* beginnen, werden unterschiedlich analysiert. Einerseits wird bei gewissen Zusammensetzungen das Lemma *willen* angegeben:

```
<token pos="152603749" len="14">
<text>Willensbildung</text>
<lemma>willen bildung</lemma>
<tags syntax="@NH" morpho="N"/>
</token>

<token pos="169858920" len="19">
<text>Willensentscheidung</text>
<lemma>willen entscheidung</lemma>
<tags syntax="@NH" morpho="N"/>
</token>
```

Andererseits das Lemma *wille*:

```
<token pos="152833708" len="12">
<text>Willenstärke</text>
<lemma>wille stärke</lemma>
<tags syntax="@NH" morpho="N"/>
</token>
```

Die neben *Willenstärke* existierende Form *Willensstärke* wird dann wiederum nach dem Muster oben analysiert:

```
<token pos="296598106" len="13">
<text>Willensstärke</text>
<lemma>willen stärke</lemma>
<tags syntax="@NH" morpho="N"/>
</token>
```

Fraglich ist, ob es, wie beim Beispiel *Embryonenforschung* oben, ebenfalls Inkonsistenzen bei der Analyse der jeweils gleichen Zusammensetzung gibt. Bei der Überprüfung von *Wissensbildung* und *Willen(s)stärke* in einem Teilkorpus scheint das jedenfalls nicht der Fall zu sein.

Auswirkung auf die Analyse:

Im Falle von *Wille-* führen die unterschiedlichen Analysen zu Verzerrungen bei der Verteilung der Zusammensetzungen zwischen *-ns* und *-s*-Schnittstellen.

Fehlerhafte Bestimmung des Erstglieds als Substantiv

Vor allem bei Zusammensetzungen mit *e*-Schnittstelle tritt das Problem auf, dass das Erstglied als Substantiv klassifiziert ist, es aber (zumindest in der Verwendung in der Zusammensetzung) keines ist. Bei der Analyse der Zusammensetzungen wurde eine Substantiv-Liste, generiert auf Basis des mit dem Xerox-Tagger getaggten DeReKo verwendet (vgl. oben). In dieser Substantivliste tauchen auch Lexeme auf, bei denen es sich offensichtlich um Fehlannotationen als Substantive handelt (Fehler 1) oder die zumindest ambig sind und in den Zusammensetzungen kaum in der Bedeutung als Substantive verwendet werden (Fehler 2). Darunter fallen Beispiele wie *Adelig*, *Lang*, *Privat* (Fehler 1) oder *Alt*, *Arm*, *Gut* (Fehler 2).

Eine genauere Analyse zeigte, dass bei den Zusammensetzungen mit *e*-Schnittstelle mit einer Fehlerrate von 5% auszugehen ist. Das hat auf die Statistiken zu den Verteilungen der Schnittstellen folgende Auswirkung:

Unkorrigierte Tabelle			Korrigierte Tabelle		
Schnittstelle	Frequenz	Frequenz %	Schnittstelle	Frequenz	Frequenz %
0	111 557 704	70,5378%	0	111 557 704	70,5545%
<i>s</i>	26 282 952	16,6187%	<i>s</i>	26 282 952	16,6226%
<i>n</i>	9 489 702	6,0003%	<i>n</i>	9 489 702	6,0017%
<i>es</i>	6 144 943	3,8854%	<i>es</i>	6 144 943	3,8864%
<i>en</i>	2 410 333	1,5241%	<i>en</i>	2 410 333	1,5244%
<i>er</i>	1 400 468	0,8855%	<i>er</i>	1 400 468	0,8857%
<i>e</i>	746 140	0,4718%	<i>e</i>	708 833	0,4483%
<i>ns</i>	41 218	0,0261%	<i>ns</i>	41 218	0,0261%
<i>ens</i>	27 274	0,0172%	<i>ens</i>	27 274	0,0172%
?	24 517	0,0155%	?	24 517	0,0155%
<i>r</i>	14 402	0,0091%	<i>r</i>	14 402	0,0091%
<i>nen</i>	5 612	0,0035%	<i>nen</i>	5 612	0,0035%
<i>st</i>	5 605	0,0035%	<i>st</i>	5 605	0,0035%
<i>ien</i>	2 142	0,0014%	<i>ien</i>	2 142	0,0014%

Tabelle II

Obwohl die Anzahl von Zusammensetzungen mit *e*-Schnittstelle von 746 140 auf etwa 708 833 (-5%) sinkt, verändert sich in der Gesamtverteilung der Schnittstellen kaum etwas. Es scheint deshalb möglich zu sein, die erhöhte Fehlerquote bei den *e*-Schnittstellen zu ignorieren.

5. Resultate

*Aus den Korpora sollten wir alles rausholen
und wir sollten möglichst nichts reinstecken.*

(Heringer 2009a, S. 31)

Die DeReKo-Schnittstellenanalysen umfassen genau 176 405 282 Schnittstellenanalysen.²⁷ Um es zu betonen: Sie umfassen nicht 176 405 282 Zusammensetzungen, sondern 176 405 282 Schnittstellen von Zusammensetzungen. Mitunter gehören nämlich zwei oder mehr Schnittstellen zu ein und der selben Zusammensetzung. Zum Beispiel analysiert der Rechner²⁸ *Barockschlossanlage* doppelt, erstens als *Barockschloss* und zweitens als *Schlossanlage*.

<i>Barockschlossanlage</i> →	1. <i>barock</i> <i>schloss</i>
	2. <i>schloss</i> <i>anlage</i>

Nach den Möglichkeiten automatischer Analyse enthält das Korpus Zusammensetzungen, die rein aus Substantiven aufgebaut sind. Mitunter irrt der Rechner dabei und mischt auch *Anhalteversuch* und *Arme-Leute-Urlaub* unter die Analysen. *Anhalteversuch* ist eine Zusammensetzung mit einem Verb als Erstglied; *Arme-Leute-Urlaub* hat eine Phrase als Erstglied. Schwierigkeiten hat der Rechner auch mit Homonymen, zum Beispiel entscheidet er sich bei *Anwaltschaft* irrtümlich für das Substantiv *Schaft* statt für das Suffixe *schaft* oder bei *Armsein* und *Festlohn* für die Substantive *Arm* und *Fest* statt für die Adjektive *arm* und *fest*. Das trübt manche Detailstatistik, ist aber angesichts der großen Datenmenge nicht wirklich relevant. Oder?²⁹

Häufig lege ich meinen Aussagen explizit die Erstglieder zugrunde, die im Korpus mehr als 10 mal vorkommen.³⁰ Dadurch will ich vor allem fehlerhafte Einzelanalysen ausschließen.

²⁷ Es sind 176 405 282 Tokens und 8 699 856 Types.

²⁸ Auf Caren Brinckmann wirkt mein Gebrauch des Wortes *Rechner*, als hielte ich Parser und Programme für magische Maschinen mit ominösem Eigenleben. Weil das tatsächlich so ist, behalte ich das Wort *Rechner* bei.

²⁹ So wird zum Beispiel die Statistik zur Schnittstelle mit *e* verunklart, weil dort auch verbale Erstglieder wie in *Hebebühne* reingerutscht sind. Dieser Rechnerfehler macht nach Berechnung von Noah Bubenhofer 5% der Schnittstellenanalysen für das Flexionselement *e* aus. Vgl. Exkurs. Muss mich das stören?

³⁰ Mit einem Vorkommen über 10 hat Noah Bubenhofer 161 541 712 Tokens und 751 216 Types errechnet.

Explizit mache ich auch Type-Token-Unterschiede. Als *Type* wird seit Charles Sanders Peirce ein Modell, eben ein Typ bezeichnet, als *Token* dessen Auftreten, dessen Vorkommen, dessen einzelne Exemplare. Im vielzitierten Satz von Ludwig Wittgenstein *Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt* gibt es einen Type *Grenze* und zwei Tokens *Grenze*.

Intuitiv werden meist Types gezählt: So geben Ortner et al. (1991, S. XXXIII) an, dass ihr Material „62 456 verschiedene [!] Substantivkomposita“ umfasst, also 62 456 Types. Aber nicht immer ist es sinnvoll, Types zu zählen. Mir geht es in dieser Studie ja um Variation. Ein Beispiel für Variation ist *Abfahrtsmöglichkeit* neben *Abfahrtsmöglichkeit*. Zähle ich hier Types, erhalte ich genau zwei und kann lediglich feststellen, dass beide Varianten zu finden sind. Zähle ich dagegen Tokens, erhalte ich 35 mal *Abfahrtsmöglichkeit* neben 14 mal *Abfahrtsmöglichkeit* und kann feststellen, dass die *s*-Variante dominiert. Je nach Anlass habe ich deshalb explizit mal Types, mal Tokens gezählt. Bei manchen Anlässen ist auch die Zählung beider im Vergleich informativ. Auch das expliziere ich jeweils. Kurz: Das Type-Token-Thema bleibt bei den Schnittstellenanalysen meiner Studie dauerpräsent.

Die DeReKo-Schnittstellenanalysen dokumentieren alle aus der Forschungsliteratur bekannten Elemente. Die meisten sind Flexionselemente, nur beim *s* kommen auch unparadigmische Elemente vor.

e

Ärzteblatt, Früchtetee, Händedruck, Hundehütte, Kräfteverhältnis, Pferdebahn, Punktestand, Schweinebraten, Städtetag, Tagebuch, Wegenetz, Zertifikate-Abteilung

*en*³¹

Abiturientenleben, Bankenaufsicht, Kandidatensuche, Menschenflut, Nachrichtensprecher, Personenkult, Satellitenfernsehen, Zeitenwende, Zwergen-Zerwürfnis

ens

*Herzensangelegenheit, Schmerzensgeld*³²

³¹ Position 1 sieht darüber hinaus eine Fuge *nen*, etwa bei Zusammensetzungen wie *Kanzlerininnenbonus*. Position 2 sieht hier den üblichen Plural *Kanzlerinnen*, dessen gedoppeltes *n* im Übrigen eine reine Schreibgewohnheit ist. Auch für Erstglieder wie *Embryonen* in *Embryonenforschung* sieht Position 1 ein Fugenelement *nen*, das an den Stamm *Embryo* angehängt wird; Position 2 sieht hier eine Pluralform *Embryonen*. Vgl. Kapitel 2.

er

Geschlechterforschung, Güterverkehr, Kinderkrippe, Kleiderbügel, Länderspiel, Männerkörper, Rinderknochen, Völkerrecht, Wörterbuch

es

Bundesregierung, Dankesrede, Freundeskreis, Kindesstatt, Landeskonferenz, Standesdünkel, Zinseszins

n

Abgabenlast, Aktienpaket, Antennenmast, Augenblick, Balladen-Nachmittag, Drachenkopf, Ehrengast, Familiengruft, Hasenpanier, Meisennest, Zangengriff

ns

*Namensrecht, Willensmensch*³³

paradigmatisches s

Altersweisheit, Anfangsverdacht, Frühlingssonne, Glückshormon, Kindsvater, Todesfahrt, Wirtshaus

unparadigmatisches s

Abfahrtsmöglichkeit, Administrationsriesen, Liebesbeweis, Nettigkeitsbeweis, Prüfungsfragen, Schönheitswettbewerb

Außerdem finden sich Zusammensetzungen mit „Subtraktionsfuge“, die durch „Subtraktion von Schwa“ entsteht (Eisenberg 2004, S. 236), etwa bei *Wolle* zu *Woll*,³⁴ sowie Zusammensetzungen mit „Nullfuge“ (ebd.). In meiner Studie spreche ich von Tilgung und Nullform.

Tilgung

Erbswurst, Kirschkuchen, Kirchgang, Kutschfahrt, Mühlrad, Sprachunterricht, Woldecke

Nullform

Airbus-Beschluss, Beinfreiheit, Fischsuppe, Frustkauf, Glascontainer, Komfortwohnung, Ringtausch, Türsturz, Zebrastreifen

³² Bei den DeReKo-Schnittstellenanalysen werden ausschließlich *Herzens* und *Schmerzens* als Formen mit *ens* analysiert. Dagegen sieht der Rechner zum Beispiel *Leidens* als *Leiden* + *s* und *Schreckens* als *Schrecken* + *s*. Er leitet sie nicht von *Leid* und *Schreck* ab, sondern von *das Leiden* und *der Schrecken*.

³³ Bei den DeReKo-Schnittstellenanalysen werden nur *Namens* und *Willens* als Formen mit *ns* analysiert. Eisenberg (2004, S. 236) nennt außerdem *Glaubens*, das der Rechner als *Glauben* + *s* analysiert. So auch *Frieden* + *s* in *Friedensmarsch*.

³⁴ Vgl. auch Fuhrhop (1996, S. 188) und Ortner et al. (1991, S. 94).

Die Verteilung dieser Wortformen und Fugenelemente zeigt sich der Tendenz nach in allen von mir berücksichtigten Untersuchungen ähnlich, im Detail aber verschieden.

Ortner et al. (1991)	%	Kürschner (2003)	%	Krott et al. (2007)	%	DeReKo- Schnitt- stellen- analysen ³⁵	%
0	72,8	0	57,6	0	65,0	0	70,5
<i>es</i>	14,8	<i>s</i> <i>es</i>	25,1 1,6	<i>s</i> <i>es</i>	17,0 1,5	<i>s</i> <i>es</i>	16,5 3,5
<i>en</i>	9,7	<i>en</i> <i>n</i>	4,2 8,9	<i>en</i>	15,0	<i>en</i> <i>n</i>	1,6 6,2
<i>e</i>	1,3	<i>e</i>	1,2	<i>e</i>	1,0	<i>e</i>	0,5
<i>er</i>	0,7	<i>er</i>	1,2	<i>er</i>	0,4	<i>er</i>	0,8
„Sonstige“	0,7	<i>nen, ens</i> u.a.	0,8	<i>ens</i>	0,2	<i>ens</i> u.a. ³⁶	0,1

Tabelle 2

Tabelle 2 zeigt, dass der gemeinsamen Tendenz nach Nullformen in der Überzahl sind; anteilig folgen das *s* und das *en*. Im Detail fällt besonders Kürschner (2003) auf. In seinem Korpus ist der Anteil der Nullformen unterdurchschnittlich, der Anteil von *s* und *er* überdurchschnittlich. Bei den DeReKo-Schnittstellenanalysen fallen der geringere Anteil von *n*, *en* und *e* auf sowie der höhere Anteil von *es*.

Weil die Korpora unterschiedlich aufgebaut und unterschiedlich groß sind, weil sich die Analyse- und Messmethoden unterscheiden, weil also Vergleiche noch mehr hinken, als sie es ohnehin schon immer tun, sagt uns der Statistikvergleich im Detail allerdings wenig. In welchem Korpus findet Sprache in ihrer Totalität denn auch statt? Und wo findet sie überhaupt statt?

³⁵ Die Berechnungen von Noah Bubenhofer beruhen auf der Auswertung von Zusammensetzungen, die mehr als 100 mal vorkommen.

³⁶ Bei „u.a.“ handelt es sich mitunter um fehlerhafte Analysen. So rechnet der Rechner ein Element *agsams* aus *Bettagsamstag* heraus, weil er annimmt, es handele sich hier um eine Zusammensetzung aus *Bett* + *Tag*. Die Analyse eines Fugen-*r* ist ebenfalls überwiegend fehlerhaft bei Zusammensetzungen des Typs *Trägerkonferenz*, bei dem der Rechner ein Erstglied *träge* + *r* annimmt. Richtig ist die *r*-Analyse bei *Aschermittwoch*; „hier geht die *r*-Form durch und man sagt nie *aschenmittwoch*“ (Deutsches Wörterbuch 1854, S. 584). So auch früher *ascherprudel*, *ascherkleid*.

Immerhin ergeben die Statistiken, dass es relativ egal ist, wie groß man den Ausschnitt innerhalb eines bestimmten Korpus wählt. Noah Bubenhofer hat mir dazu eine Statistik erstellt, in der er Teilkorpora nach Größe abgestuft hat. Für die drei häufigsten Erstgliedformen ergeben sich dabei ähnliche Anteile. Gerechnet wurden Tokens.

	Teilkorpus I	Teilkorpus II	Teilkorpus III	Teilkorpus IV
Zahl der Zusammensetzungen	176 405 282	68 296 495	23 495 386	12 227 778
Anteil der Zusammensetzungen am Gesamtkorpus	5,1%	5,2%	4,8%	5,1%
0	70,5%	70,8%	69,3%	69,8%
<i>s</i>	16,5%	16,2%	17,5%	17,4%
<i>n</i>	6,2%	6,2%	5,8%	6,0%

Tabelle 3

5.1 Regeln

Linguisten vermuten immer mal wieder die eine oder andere Regel, nach der wir Zusammensetzungen im Schnittstellenbereich gestalten. Meine Idee für diese experimentelle Studie ist, solche Regeln durchzuspielen, denn meine Studie dokumentiert ja Variation und wir variieren vor allem da, wo sich Regeln widerstreiten. Die Regeln und ihre Befolgung oder Nichtbefolgung geben also meiner Studie Struktur. Ich habe sie auf fünf Nenner gebracht:

- Paradigmenregel
- Bedeutungsregel
- Euphonieregel
- Strukturierungsregel
- Gewohnheitsregel

Paradigmenregel

**Nach der Paradigmenregel
bilden wir Zusammensetzungen mit
paradigmischen Formen von Erstgliedern.**

Das Erstglied ist überwiegend eine Form seines Flexionsparadigmas. Das lässt sich schnell rausrechnen: Unparadigmische Erstglieder sind ausschließlich Erstglieder mit dem Fugen-*s* (*Kindheits*), das heißt Erstglieder, in deren Flexionsparadigma keine *s*-Form existiert, die sich in Zusammensetzungen aber mit dem *s* verbinden. Alle anderen Erstglieder haben paradigmische Formen.³⁷ Gerechnet hat Noah Bubenhofer wiederum Tokens, die mehr als 10 mal vorkommen.

paradigmische Erstglieder	unparadigmische Erstglieder
155 400 821	14 301 478

Tabelle 4

Tabelle 4 zeigt, dass nur 8,4% der Erstglieder in unparadigmischer Form vorkommen, nämlich mit dem unparadigmischen *s*,³⁸ und 91,5% der Erstglieder paradigmisch sind. Das ist eine gute Nachricht, denn Lerner haben kaum Überraschungen zu erwarten. Niemand muss zum Beispiel mit *Katzerkorb* oder *Kindenzimmer* rechnen. Die Regel der Paradigmentreue ist relativ zuverlässig.

Aber welche paradigmischen Formen verwenden wir im Einzelfall? Hier kommt die schlechte Nachricht: Linguisten tun sich mit Antworten schwer – schon weil deutsche Flexionsformen oft nicht eindeutig zeigen, was sie sind und was nicht. Mit was haben wir es zum Beispiel bei *Krakenarme* zu tun? Werden die Arme eines Kraken bezeichnet oder die Arme vieler Kraken? Bezeichnet *Lehrerzimmer* ein Zimmer für den Lehrer oder ein Zimmer für die Lehrer? Bezeichnet *Botennachricht* die Nachricht eines Boten oder die Nachricht mehrerer Boten?

³⁷ Wiederum mit den Ausnahmen der Tilgungsformen wie *Woll* in *Wollknäuel*.

³⁸ Der Anteil des unparadigmischen *s* an den *s*-Formen ist allerdings relativ hoch: Nach einer Berechnung von Noah Bubenhofer sind Erstglieder mit paradigmischem Genitiv-*s* (*Königs*) zu 49,3% an den Schnittstellenanalysen mit *s* beteiligt, Erstglieder ohne paradigmisches *s* (*Kindheits*) mit 50,6%. Gut die Hälfte der *s*-Formen sind also unparadigmisch. Vgl. zur Berechnung der paradigmischen und unparadigmischen Formen auch den Exkurs.

Eindeutig immerhin sind viele Genitive wie *Lehrers* in *Lehrerstochter* oder *Herzens* in *Herzensbildung*, denn *Lehrers* und *Herzens* kommen in ihren Paradigmen allein als singularische Genitivform vor: *des Lehrers*, *des Herzens*. Auch *Dokumenten* in *Dokumentenmappe* kommt im Paradigma von *Dokument* allein als Dativ Plural vor: *den Dokumenten*.

Diskutiert wird in der Linguistik außerdem, ob *s*-Plurale als Erstglieder verwendet werden. Sie machen nach Marcus et al. (1995, S. 239f.) insgesamt 4% aller deutschen Plurale aus. Den *s*-Plural gibt es in allen drei Genera, im Maskulinum, im Neutrum und im Femininum.

<i>der Clown</i>	<i>das Auto</i>	<i>die Bar</i>
<i>die Clowns</i>	<i>die Autos</i>	<i>die Bars</i>

Wegener (2003, S. 433) nennt maskuline Erstglieder wie in *Clownsparade*. Offenbar ist für sie pluralische Interpretation ausschlaggebend, „obwohl das Erstglied dann teilweise auch als Genitiv Singular-Phrase interpretierbar ist“ (ebd., S. 435). Dasselbe gilt für neutrale Erstglieder wie *Training*, *Bonbon* oder *Gel*³⁹ und für den in „colloquial speech, specifically in areas of northern Germany“ (Fehringer 2009, S. 156) üblichen maskulinen und neutralen Plurale *die Mädels*, *die Jungs*, *die Kerls*, *die Kumpels*, *die Onkels*, die auch im Singular ein *s* haben: *des Mädels*, *des Onkels*. Eindeutige Pluralformen sind Feminina wie *Bars*, *Kameras*, *Pizzas*, die im Singular kein *s* haben: *Besitzer der Bar*. Ausgerechnet diese einzigen eindeutigen femininen *s*-Plurale werden nun aber gerade nicht als pluralische Erstglieder realisiert, denn es heißt *Barhocker*, *Kameraeinstellung* und *Pizzaservice*, nicht *Barshocker*, *Kameras-einstellung* und *Pizzasservice*.

Bedeutungsregel

**Nach der Bedeutungsregel
wählen wir Erstglieder entsprechend ihrer Bedeutung.**

Im Gegensatz zu Satzstrukturen ist „bei den Wortbildungen gerade das Fehlen von expliziten Mitteln kennzeichnend, welche die [...] Relationen verdeutlichen würden“ (Eichinger 2000, S. 51). Tatsächlich gilt bei mindestens 70% der DeReKo-Schnittstellenanalysen das Prinzip der Bedeutungs vagheit. In

³⁹ Vgl. auch Elshorbagy (2007, S. 9).

über 70% deutscher Zusammensetzungen stehen Wörter relationslos in der vagen Nullform nebeneinander, zum Beispiel in *Abendsonne* und *Architekturmuseum*.

Welche Bedeutung gerade aktualisiert wird, können wir solchen Zusammensetzungen nicht von ihrer Gestalt ablesen. Da ist nämlich „nichts, kein Zeichen für diese Relation. [...] Und war es nicht ein gutes altes strukturalistisches Prinzip: Wo kein Ausdruck, auch keine Bedeutung?“ (Heringer 1984, S. 3). Bei Zusammensetzungen des Typs *Abendsonne* kann zwar die Reihenfolge von *Abend* und *Sonne* Aufschluss geben, „entweder A vor B oder B vor A“, aber: „Das ist alles.“ (ebd.).⁴⁰

Gelegentlich möchten wir aber doch einer Zusammensetzung auch äußerlich etwas mehr Bedeutung mitgeben. Zum Beispiel möchten wir sichtbar Vielzahl betonen. Wir sagen *Ärztehaus*, um zu betonen, dass dieses Haus von mehreren Ärzten betrieben wird. Wir sagen *Länderkonferenz* und meinen oder verstehen eine Konferenz, an der pluralisch Länder beteiligt sind. Sprachenlehrer vermitteln explizit mehrere Sprachen. Und bei einem Burgen- und Schlösserradweg dürfen wir darauf vertrauen, an mehr als einer Burg und mehr als einem Schloss vorbeiradeln zu können.

⁴⁰ Willems (2001, S. 150) nennt das eine „syntaktische Nullstruktur“. Demnach wird auch keine Relation in den sogenannten Rektionskomposita angezeigt, bei denen aber nach Forschungsmeinung „eine morphologisch strukturelle Präferenz für die Bedeutungszuweisung besteht“ (Eisenberg 2004, S. 230). Als Rektionskomposita werden Zusammensetzungen gesehen mit einer bestimmten Relation zwischen Erst- und Zweitglied, nämlich der rektionalen Relation. Typische Rektionskomposita sind *Romanleser*, *Konfliktbewältigung* und *Wetterbeobachtung*. Unter Rektion (zu lateinisch *regere* ‘leiten, beherrschen’) verstehen Linguisten die syntaktische Abhängigkeit eines Wortes von einem anderen: So regiert *kennen* den Akkusativ, indem *kennen* steuert, dass sein Objekt im Akkusativ steht: *er liest den Roman*. Werden Verben zu Substantiven abgeleitet, vererben sie diese Rektion an das Substantiv: *Leser des Romans*. Rektionskomposita wie *Romanleser* bestehen aus einem solchen semantisch sättigungsfähigen, sättigungsbedürftigen Substantiv als Zweitglied (*Leser*) und einem Erstglied, das den Argumenten der zugrundeliegenden Verben entspricht (*Roman*). Die Argumentstruktur bleibt im Wesentlichen erhalten. Das aber sehen wir nicht an der Gestalt solcher Zusammensetzungen; wir erschließen es vielmehr aus unserem Kontext- und Weltwissen.

Vielzahl können wir durch Pluralmarker betonen.⁴¹ Als Pluralmarker kommen für die Erstglieder deutscher Zusammensetzungen *en*, *n*, *e*, *er* und *nen* in Frage. Diese Plurale machen knapp 9% aller Erstglieder aus. Die Belege habe ich aus den DeReKo-Schnittstellenanalysen zu illustratorischen Zwecken ausgewählt. Sie sind typisch, aber nicht unbedingt besonders häufig. Gerechnet habe ich Tokens.

0	70,5%
<i>n</i> <i>Abgabenlast, Anzeigenfeld</i>	6,0%
<i>en</i> <i>Artenschutz, Autorenschaft</i>	1,5%
<i>er</i> <i>Ämterfrage, Bäderangebot</i>	0,8%
<i>e</i> <i>Ärztenschwemme, Gästehaus</i>	0,4%
<i>ens</i> und andere <i>Herzenswunsch, Schmerzensschrei</i>	0,1%

Tabelle 5

Aber auch mit der Verwendung von Genitivformen kann ein Mehr an Bedeutung herausgeschlagen werden. Deutsche Zusammensetzungen gehen historisch aus Genitivphrasen wie *gotes boto* oder *wintes brut* zurück. Der genitivische Bezug spricht immer noch aus *Gottesbote* und *Windsbraut*. Und wir können ihn auch Zusammensetzungen wie *Landesvater* oder *Bäckerstochter* mitgeben.⁴² Als überwiegend eindeutige Genitivmarkierer kommen für die

⁴¹ Ich meine damit ausdrücklich: können! Denn es gilt: Überall, wo wir in Zusammensetzungen Vielzahl betonen wollen, wählen wir ein Erstglied mit Pluralmarker. Aber nicht überall, wo ein Erstglied mit Pluralmarker gewählt wird, soll Vielzahl betont werden. Die Bedeutungsregel ist wie alle anderen Regeln ein- wie aussetzbar.

⁴² Vgl. zu impliziten Genitivsignalen Lauterbach (1993, S. 204-221).

Erstglieder in deutschen Zusammensetzungen vor allem *es* und *s* in Frage.⁴³ Die authentischen Belege habe ich wiederum aus den DeReKo-Schnittstellenanalysen ausgewählt; gerechnet habe ich wiederum Tokens.

0	70,4%
paradigmatisches s <i>Auftragslage, Lehrersfrau</i>	16,8%
es <i>Diebestrio, Tageslicht</i>	3,6%

Tabelle 6

Euphonieregel

**Nach der Euphonieregel
lassen wir Zusammensetzungen schön klingen.**

Euphonie (zu griechisch *eu* ‘schön’) ist der schöne Klang, der Wohlklang im weitesten Sinne. Aber was klingt schön?

„Immer wieder liest man, dass Fugenelemente der Artikulationserleichterung dienen sollen. Tatsächlich aber führt gerade das frequenteste Element *-s-* immer zu einer Komplexitätssteigerung der Silbenkoda, häufig sogar zu extrasilbischen Einheiten: *Mehrheit+s+meinung*, *Wissenschaft+s+markt*“ (Nübling/Szczepaniak 2009, S. 202).⁴⁴ Im Übrigen tun sich deutsche Sprecher ohnehin nicht besonders schwer mit der Artikulation abenteuerlicher Konsonantenfolgen und holpernder Extrasilben. Problemlos sagen sie *Angst*, *Vernunft* und *Landsknecht*.

Fraglich sind insofern auch Thesen zum Nichtpassen bestimmter Lautfolgen im Schnittstellenbereich. Zum Beispiel „scheint“ für Ortner et al. (1991, S. 110) „nach dem Befund des Korpus und der überprüften Wörterbücher der

⁴³ Weniger eindeutig sind andere Flexionselemente wie *n* in *Bauernfrühstück*: Bezeichnet *Bauernfrühstück* das Frühstück des Bauern oder der Bauern? Dass auch *s* nicht immer eindeutig ist, ist oben bei der Paradigmenregel zur Sprache gekommen: Mitunter gibt es *s* in ein und demselben Flexionsparadigma als Singular-Genitiv- und als Pluralmarker, zum Beispiel bei singularem *des Autos* und pluralem *die Autos, der Autos, den Autos*.

⁴⁴ Vgl. schon Nübling/Szczepaniak (2008, S. 6).

Auslaut [f] selten ein Fugenzeichen zuzulassen: *Briefabschrift, Topfflicker, Motivforschung*. Ausgenommen sind die Maskulina mit der Pluralendung *-en* (*Nervenzelle, Grafenwürde*), die mehrsilbigen maskulinen Verbalsubstantive (*Bedarfsfall*) und eine kleine Gruppe von Wörtern, bei der die *-s*-Fuge stammgebunden ist (*Maulwurf, Wolf, Bischof, Friedhof, Bahnhof*)“ (Ortner et al. 1991, S. 110). Auch in den DeReKo-Schnittstellenanalysen tendieren maskuline Mehrsilber auf [f], die aus Verben abgeleitet sind, zum Genitiv-*s*, zum Beispiel in *Angriffsflug, Ausverkaufsfieber, Bedarfsfall* und *Vorwurfsform*. Und es kommen einige andere genitivische Erstglieder vor wie in *Detektivs-outfit, Schafsmilch* und *Schiffsanker*. Dass Erstglieder wie *Detektivs* „eine kleine Gruppe von Wörtern“ sind, liegt allerdings einfach daran, dass sie überhaupt eine kleine Gruppe sind, zumal im Vergleich zu der großen Gruppe der aus Verben abgeleiteten Substantive. Und ist „selten“ adäquat resümiert? Von den 27 118 084 Schnittstellenanalysen mit *s* haben immerhin 593 364 die Abfolge ‘Erstglied auf *f* plus *s*’. Hier greift vielleicht einfach, was Perkuhn/Belica (2006, S. 5) resümieren: „Je größer die Datensammlungen sind, desto mehr seltene Phänomene decken sie mit ausreichender Aussagekraft ab.“⁴⁵

Schauen wir uns noch eine andere lautliche Vereinbarkeitsthese an: Elshorbagy (2007, S. 19) beobachtet, dass *s* besonders Erstgliedern auf *t* folge, also Erstgliedern wie *Arbeit, Heirat, Fastnacht, Weihnacht, Auskunft, Andacht, Einfalt*. Meine Recherchen ergaben, dass bei den DeReKo-Schnittstellenanalysen tatsächlich Erstglieder auf *t* häufiger ein *s* bekommen als andere konsonantisch endenden Erstglieder. Gezählt habe ich wiederum Tokens und zwar solche, die mehr als 10 mal vorkommen.

Erstglieder auf <i>t</i> + <i>s</i>	9 660 823
Erstglieder auf <i>g</i> + <i>s</i>	8 633 181
Erstglieder auf <i>n</i> + <i>s</i>	3 476 928
Erstglieder auf <i>d</i> + <i>s</i>	1 042 304
Erstglieder auf <i>r</i> + <i>s</i>	832 650
Erstglieder auf <i>f</i> + <i>s</i>	593 364
Erstglieder auf <i>l</i> + <i>s</i>	438 906

Tabelle 7

⁴⁵ Sie schließen übrigens daraus, dass „erst ab einer gewissen Größenordnung Korpora mehr Wissen als die Summe ihrer Belege“ enthalten.

Allerdings vermutet Wegener (2005, S. 181), dass „das schon von Augst 1975 beobachtete besonders häufige Auftreten von *-s* nach finalem *-t* ein Zufallsprodukt, bedingt durch die hohe Zahl von auf *-t* auslautenden Stämmen“ ist. Um das zu prüfen, habe ich die konsonantischen Erstglieder der Tabelle 7 insgesamt ermittelt und dann geschaut, wieviel Prozent davon *s*-Formen sind. Gerechnet habe ich wieder Erstglieder, die mehr als 10 mal vorkommen, und wiederum Tokens. Für diese Frage ist nämlich die Zählung der Tokens informativ, denn es geht ja gerade um häufiges Auftreten. Die Erstglieder habe ich nach ihrer Auftretenshäufigkeit angeordnet.

Erstglieder auf <i>t</i> insgesamt	86 718 438	davon <i>s</i> -Formen	33,4%
Erstglieder auf <i>r</i> insgesamt	70 171 686	davon <i>s</i> -Formen	3,5%
Erstglieder auf <i>n</i> insgesamt	44 035 407	davon <i>s</i> -Formen	23,6%
Erstglieder auf <i>d</i> insgesamt	40 302 825	davon <i>s</i> -Formen	7,5%
Erstglieder auf <i>g</i> insgesamt	36 769 581	davon <i>s</i> -Formen	70,4%
Erstglieder auf <i>l</i> insgesamt	34 990 653	davon <i>s</i> -Formen	3,7%
Erstglieder auf <i>f</i> insgesamt	8 822 814	davon <i>s</i> -Formen	20,1%

Tabelle 8⁴⁶

Tabelle 8 zeigt erstens, dass Erstglieder auf *t* tatsächlich extrem häufig sind. Darunter sind 2 321 748 Erstglieder auf *heit* und 511 212 Erstglieder auf *keit*, die zu großen Teilen das unparadigmische *s* verlangen: Es haben 18 141 Erstglieder auf *heit* und 2 319 Erstglieder auf *keit* die Variante mit *en*,⁴⁷ während sich 2 294 787 Erstglieder auf *heit* und 507 924 Erstglieder auf *keit* mit *s* verbinden. Zweitens bestätigt die Tabelle 8 Wegeners Vermutung, denn prozentual verbinden sich nicht die Erstglieder auf *t*, sondern die deutlich selteneren Erstglieder auf *g* am häufigsten mit *s*, darunter 19 144 605 auf *ung*, aber auch Erstglieder wie *Alltag* und *Samstag*.

Als euphonisches Prinzip favorisieren Linguisten außerdem die „rhythmische Optimierung“ (Nübling/Szczepaniak 2009, S. 203). Nach Wegener (2003, S. 453) mögen deutsche Zungen und Ohren besonders den Trochäus. Nüb-

⁴⁶ Recherchiert habe ich Erstglieder auf grafisches *f*. Hierzu könnte man die 873 384 Erstglieder auf *v* zählen, unter denen sich allerdings nur 21 mit *s* verbinden (*Detektivsoutfit*). 724 197 haben die Nullform, 109 881 die *en*-Form.

⁴⁷ Rechnungen mit mehr als 10 mal vorkommenden Erstgliedern ergeben hier 56 Types auf *heit* plus *s* und 148 Types auf *keit* plus *s*.

ling/Szczepaniak (2009, S. 207) sehen im Trochäus „das phonologische Wortideal“. Trochäische Erstglieder bestehen aus einer schweren betonten und einer leichten unbetonten Silbe: *Hünde*.⁴⁸ Nach dem Euphonieprinzip bevorzugen wir trochäische Erstglieder wie *Hunde*, *Hühner* und *Kinder* in *Hundehütte*, *Hühnerstall* und *Kindergarten*.

Auch dass Erstglieder auf *heit*, *ion*, *ität*, *keit*, *ling*, *sal*, *schaft*, *tum* und *ung* regelmäßig⁴⁹ ein *s* erhalten, etwa bei *Kapazitätswuchs* und *Schicksalsort*, begründen Wegener (2003, S. 445) und Nübling/Szczepaniak (2009, S. 220) mit dem Euphonieprinzip: „Die Fuge dient aus wortphonologischer Perspektive der Anzeige beziehungsweise Reparatur schlechter, das heißt auffällig strukturierter phonologischer Wörter“ (ebd.). Hier Belege aus den DeReKo-Schnittstellenanalysen:

heit⁵⁰

Abwesenheitsnotiz

ion

Abstraktions-Ansatz

ität

Identitätsklau

keit

Aufmüpfigkeitspotenzial

ling

Schmetterlingsdekor

schaft

Wissenschaftskolleg

tum

Altertumsfimmel

ung

Überzeugungshansestädter

⁴⁸ Das Silbengewicht richtet sich nach der Dauer einer Silbe. Offene Silben mit Kurzvokal sind von kurzer Dauer; sie werden *leichte Silben* genannt. Längere Silben werden *schwere Silben* genannt.

⁴⁹ Regelbestätigende Ausnahmen sind Pluralvarianten wie in *Minderheitenregierung* und *Antiquitätenhandel*. Vgl. Fuhrhop (2000, S. 203).

⁵⁰ Eine Ausnahme ist *Minderheit* mit der Variante *Minderheiten*. Vgl. Fuhrhop (1996, S. 541).

Bei den DeReKo-Schnittstellenanalysen haben von den 81 354 252 Erstgliedern mit *s* 9 114 855 Erstglieder diese Endungen, also 11,2%. Das häufige Vorkommen liegt natürlich auch am häufigen Vorkommen solcher Erstglieder. Gerechnet habe ich wiederum Erstglieder, die mehr als 10 mal vorkommen. Die Zahlen informieren vor allem über die Häufigkeit des Auftretens. Das macht der Vergleich der Tokens mit den Types deutlich. Manche Erstglieder treten extrem häufig auf. So entfallen in Tabelle 9 zum Beispiel von den 4 004 409 Tokens auf *schaft* 1 911 831 auf *Wirtschaft*, 239 331 auf *Mannschaft* und 65 643 auf *Bereitschaft*. Bei Erstgliedern auf *sal* haben wir es mit genau einem Type zu tun, nämlich mit *Schicksal*, das aber 28 275 mal vorkommt.

	Tokens	Types
Erstglieder auf <i>ung</i> + <i>s</i>	19 144 605	1 347
Erstglieder auf <i>ion</i> + <i>s</i>	5 337 402	384
Erstglieder auf <i>schaft</i> + <i>s</i>	4 004 409	52
Erstglieder auf <i>heit</i> + <i>s</i>	2 294 787	56
Erstglieder auf <i>keit</i> + <i>s</i>	507 924	148
Erstglieder auf <i>ling</i> + <i>s</i>	505 599	26
Erstglieder auf <i>tum</i> + <i>s</i>	199 911	13
Erstglieder auf <i>sal</i> + <i>s</i>	28 275	1

Tabelle 9

Schließlich kann, wie Fehringer (i.Dr., o.S.) Wegener referiert, das *s* „nach schweren Silben am Wortende auftreten, z.B. *Autos*, nicht jedoch nach schweren Silben innerhalb eines Kompositums, z.B. **Autosbahn*. [...] Ein Fugenelement ist demnach zulässig in Komposita wie *Kukuck-s-ei*, jedoch nicht in Komposita wie **Uhu-s-ei*, da in diesem Fall das Erstglied auf eine schwere Silbe endet“. Fehringer (ebd.) sieht das allerdings skeptisch, denn es konnten „phonologische Argumente, welche als alternative Erklärung für das Nichtvorkommen von *-s-* in bestimmten Komposita im Standarddeutschen angebracht wurden, nicht erfolgreich auf die schwäbischen Daten angewendet werden.“⁵¹

⁵¹ Vgl. schon Fehringer (2009).

Strukturierungsregel

**Nach der Strukturierungsregel
strukturieren wir Zusammensetzungen überschaubar.**

Deutsche Zusammensetzungen sind zwar nicht so unüberschaubar wie ihr weltweiter Ruf verheißt,⁵² aber auch bei weniger langen Zusammensetzungen haben wir mitunter das Bedürfnis, dem Hörerleser „eine Dekodierungshilfe zu bieten“ (Nübling/Szczepaniak 2009, S. 204). Nübling/Szczepaniak illustrieren das an Zusammensetzungspaaren.

Hofbeamter – Bahnhofsbeamter

Nach der Strukturierungsregel verwenden wir bei auffällig komplexen Zusammensetzungen als Erstglieder eine auffällige Form, die die Soll-Schnittstelle hervorhebt. Zur Erinnerung: Typische deutsche Zusammensetzungen sind immer zweigeteilt; immer sind sie aus einem untergeordneten Erst- und einem dominanten Zweitglied aufgebaut. Bei höher komplexen Zusammensetzungen ist aber nicht gleich ersichtlich, wo die Schnittstelle sein soll, was zum Erst- und was zum Zweitglied gehört. Wortformen und Fugenelemente sollen hier strukturieren helfen. Ganz sicher sind sich Linguisten da aber nicht. Strittig ist auch die Strukturierungsabsicht des Bindestrichs, typografisch Viertelgeviertstrich genannt. Er ist eine rein grafische Hilfe, also eine Nachricht des Schreibers an seinen Leser. Die amtliche Orthografieregeln, die den Bindestrich bei Zusammensetzungen aus Substantiven regelt, ist der § 45. Danach kann der Bindestrich gesetzt werden zur Gliederung unübersichtlicher Zusammensetzungen wie bei *Arbeiter-Unfallversicherungsgesetz*, zur Vermeidung von Missverständnissen wie bei *Drucker-Zeugnis* versus *Druck-Erzeugnis* oder beim Zusammentreffen von drei gleichen Buchstaben wie bei *Kaffee-Ersatz*.⁵³

Hyvärinen (1997, S. 311-315) sieht für das Setzen des Bindestrichs oft keinen „erkennbaren Grund, denn auch semantisch normal erscheinende Verbindungen werden dekomponiert“. Als Beispiele nennt sie Botho Strauß' *Finger-*

⁵² So findet etwa der deutschlandreisende Mark Twain (1876, S. 43): „Diese langen Dinger sind kaum echte Wörter, sondern eher Wortkombinationen, und ihr Erfinder hätte umgebracht werden müssen.“ Als Beispiel nennt er das – für unsere deutschen Augen ja noch recht überschaubare – *Generalstaatsverordnetenversammlung*. Dagegen weist Augst (2001) nach, dass lange Zusammensetzungen – anders als von Sprachpflegern beklagt – keineswegs im Übermaß und auch heute keineswegs häufiger als früher gebildet werden.

⁵³ Vgl. <http://www.rechtschreibrat.com/rechtschreibung/regelwerk.html>.

Weisheit, Kälte-Anwurf und *Alb-Nacht*. „Zum Teil sind Strauß' Bindestrichkomposita wie zufällige Schnappschüsse, in denen der Strich den okkasionellen Charakter hervorhebt“ (ebd., 313).⁵⁴ Insofern sieht sie in Bindestrichen vor allem ein Zeichen für Okkasionalität. So sieht es auch Satkauskaitė (2008, S. 254), deren Korpusanalyse ergibt, dass Zusammenschreibung überwiegt, je „mehr die Bestandteile eines Kompositums konventionalisiert sind“ und „je mehr das Kompositum selbst konventionalisiert ist“.

Kauffer (1993, S. 432) hält dagegen die strukturierende Funktion des Bindestrichs für „bemerkenswert, denn er kommt an präzise festgelegten Stellen der zwei- und mehrgliedrigen Nominalkomposita vor: zwischen den UK,⁵⁵ in der ersten, aber nie innerhalb der zweiten UK, was auf pragmatische Ziele zurückzuführen ist“. Der Strukturierungseffekt wird besonders deutlich bei höher komplexen Zusammensetzungen. Dazu habe ich allenthalben Belege gefunden.

Clack

Eierschalen-Sollbruchstellen-Verursacher

Edelstahl poliert

(Ikarus design katalog 2010, S. 118)

*nach hinten zu den blitzenden **Edelstahl-Softemaschinen***

(Yglesias 2002, S. 65)

*Ich war wütend auf meinen Vater, weil er mich aus dem Bett gezerrt und der **Neonröhren-Busfahrt** und seinem zweifelhaften Körpergeruch ausgesetzt hatte.*

(Toutonghi 2009, S. 67)

Vielfach ist zu lesen, dass sich besonders das Fugen-s nicht mit dem Bindestrich vertrage, „also nicht **Geburtstags-Rabatte*, sondern *Geburtstagsrabatte*“ (Nübling/Szczepaniak 2009, S. 206). Dagegen spricht die Beleglage. Problemlos habe ich sowohl paradigmatisches als auch unparadigmatisches s plus Bindestrich sammeln können. Die gefundenen Schreibungen zeigen, dass das s als zum Erstglied gehörig empfunden wird.

*Joseph Roth spricht auch von der **Bahnhofs-Romantik**.*

(Frankfurter Allgemeine Sonntagszeitung, 30.9.2007, S. 16)

*Der Fall war im Februar 2008 aufgefliegen, nachdem drei Kinder die **Missbrauchs-Vorfälle** einer Mutter erzählt hatten.*

(Rhein-Zeitung 2008, IDS-Korpora)

⁵⁴ Vgl. auch Ruge (2006, S. 475).

⁵⁵ UK steht für *Unmittelbare Konstituente* und meint hier die Glieder einer Zusammensetzung. Zu Terminus und Begriff vgl. Fleischer/Barz (1995, S. 43).

das Leben eines Möchtergern-Universitäts-Absolventen

(Curtin 2009, S. 192)

Der Erkältungs-Doktor hilft über Nacht.

(Printwerbung für Melissegeist, 2009)

Auch die DeReKo-Schnittstellenanalysen enthalten zahlreiche Verbindungen von Wortformen beziehungsweise Fugen-s mit dem Bindestrich.⁵⁶ Recherchiert habe ich wiederum nur Erstglieder, die mehr als 10 mal vorkommen. Die Belege in Tabelle 10 sind wiederum typisch, aber nicht unbedingt besonders häufig.

0 + Bindestrich <i>Blechbläser-Musik</i> <i>Orient-Küche</i>	25 879 203
s + Bindestrich <i>Podiums-Diskussion</i> <i>Verbands-Spruchkammer</i>	1 780 107
n + Bindestrich <i>Apotheken-Notdienst</i> <i>Listen-Chef</i>	1 052 124
en + Bindestrich <i>Graduierten-Kolleg</i> <i>Patienten-Forum</i>	651 129
es + Bindestrich <i>Gesetzes-Dschungel</i> <i>Tages-Tourismus</i>	642 624
e + Bindestrich <i>Angelfreunde-Team</i> <i>Gäste-Coach</i>	271 437
er + Bindestrich <i>Kinder-Zaubershow</i> <i>Liebeslieder-Walzer</i>	170 259

Tabelle 10

⁵⁶ Hier ist die Suche so eingerichtet, dass alle Zusammensetzungen erfasst werden, in denen der Rechner an beliebiger Schnittstelle das Flexions- beziehungsweise Fugenelement gefunden hat. Zum Beispiel sind beim Genitiv-es auch Zusammensetzungen subsummiert wie *Bundesliga-Tabelle* oder *Kajak-Bundestrainer*.

Die DeReKo-Schnittstellenanalysen erlauben keine Suche nach Zusammensetzungen mit strukturierendem Leerzeichen oder strukturierender Binnengroßschreibung.⁵⁷ Vermutlich finden sich aber ohnehin kaum Belege dafür ins den IDS-Korpora, weil Leerzeichen und Binnengroßschreibung uns überwiegend im wahren Leben begegnen. Jedenfalls habe ich sie am ehesten auf Produktzetteln oder Schaufensterbeschriftungen gefunden.⁵⁸ Auch hier kann man Strukturierungsabsichten annehmen. Oder auf englische und computersprachliche Einflüsse schimpfen.

extra **Pflaumen Mus**

(Aufschrift auf Schwartau extra 2007)

Die coolste Agentin der **BuchWelt**

(www.dtv.de, August 2010)

Mit dem neuen **HausarztProgramm** *spart sich der Allgemeinmediziner Bürokratie und gibt den AOK-Versicherten ein Mehr an Leistung.*

(AOK-Life 5/2008, S. 12)

Gewohnheitsregel

**Nach der Gewohnheitsregel
nehmen wir gerne etablierte Zusammensetzungen
als Vorbilder.**

Wie wir heute Substantive zusammensetzen, „richtet sich in hohem Maß nach lexikalisch fixierten Mustern“ (Plank 1981, S. 39).⁵⁹ Wir nehmen vertraute, etablierte Zusammensetzungen gerne als Vorbild. Wer *Herrentorte*, *Herrenrad* und *Herrenabend* tausendmal gehört hat, der wird *Herrenkimono* und *Herrenknie* in Ordnung finden. Wer *Schwanenhals* kennt, bildet *Schwanengebrumme* (taz 1997, IDS-Korpora) oder *Schwanenwindbeutel* (Rheinische Zeitung 2002, IDS-Korpora), auch ohne zu wissen, dass *Schwanen* ein historischer Plural ist.

⁵⁷ Dürscheid (2006, S. 159), schließt nicht aus, dass strukturierende Binnengroßschreibung zur Normalität wird.

⁵⁸ Das ist übrigens ein gutes Argument für die Erweiterung unserer Korpora um genau solche Texte, die den Schreibstil der meisten Menschen mehr beeinflussen als belletristische oder Zeitungstexte.

⁵⁹ Vielleicht meint Furhhop (1998) mit ihren „Kompositionsstammformen“ auch die Gewohnheitsregel. Ihrer Auffassung nach haben ja Erstglieder zusammensetzungsspezifische Formen, mit denen wir immer wieder auf Wortbildungssituationen reagieren.

*der **schwanen** sterbelied,
was star und älster schwatzen.*

(Hagedorn o.J., Deutsches Wörterbuch 1899, S. 2202)

In Zusammensetzungen konservieren wir überhaupt historische Formen wie den nicht mehr gebräuchlichen Genitiv Singular *des Sonnen*⁶⁰ oder den auch nicht mehr gebräuchlichen Genitiv Plural *der Sternen*.⁶¹

*recht also ganz ein luter glas
belibet, das
des sunnen schin durchglenzet.*

(Handschrift 1372, Deutsches Wörterbuch 1941, S. 1591)

*es ist kein kraut auf erden,
das nicht ein stern im himmel hab etc.,
der zu ihme sagt: wachse.
so schießt nun alle krafft der **sternen**
auf die erden.*

(Prätorius 1669, Deutsches Wörterbuch 1941, S. 2460)

Auch greifen wir wohl aus Gewohnheit bei allen Ableitungen auf *heit*, *ion*, *ität*, *keit*, *schaft* oder *ung* zum unparadigmatischen Fugen-*s*.

5.2 Vorkommen und Verteilung von Varianten

Vermutlich ergeben sich die Erstgliedvarianten aus dem Widerstreit der Regeln. So stellt etwa Wegener (2005, S. 183f.) fest: „Hat nun eine Stammsilbe selbst schon einen oder – horribile dictu – gar zwei extrasilbische Konsonanten“, wird die *s*-Form vermieden, zum Beispiel bei *Markttag*, „allerdings kann diese Restriktion dann überspielt werden, wenn ein komplexes Kompositum vorliegt“, zum Beispiel bei *Jahrmarktsbude*. Hier widerstreiten Euphonie- und Strukturierungsregel. Hartkamp/Schneider-Wiejowski (2010, S. 209) schließen aus ihren Neologismendaten, dass sich ihre Regel 1, nach der das *s* „an den Auslaut *[-n]* gebunden ist“, gegenüber ihrer Regel 11 durchsetzt, nach der Kontinuitiva und Singulariatantum immer eine Nullfuge haben. Hier widerstreiten Euphonie- und Bedeutungsregel. Und schon Plank (1981, S. 39)

⁶⁰ Zu im Mittelhochdeutschen verbreitetem maskulinem Nominativ. Heute nur noch als Femininum *die Sonne* mit dem Genitiv Singular *der Sonne*.

⁶¹ Zu schwachem Maskulinum, analog noch heutigem *der Boten*, *der Drachen*, *der Löwen*. Heute nur noch als starkes Maskulinum *der Sterne*, analog *der Körbe*, *der Söhne*, *der Ströme*.

sieht einen „Widerstreit zweier Regularitätskonzeptionen: Orientierung an grammatischen Regeln (Wahl des Fugenelements geregelt durch synchronisch gültige Flexionsklassenzugehörigkeit) versus Orientierung an lexikalischen Mustern (die historische Flexionsverhältnisse tradieren)“. Hier widerstreiten Paradigmen- und Gewohnheitsregel.

Gibt es ein Ranking der Regeln? Ist zum Beispiel die oberste Regel die Paradigmenregel, weshalb wir zu 91,5% Erstglieder in einer Form wählen, die seinem Flexionsparadigma entspricht? Und wo überall spielt die Macht der Gewohnheit mit hinein? Die Regeln widerstreiten sich nämlich nicht nur, sie ziehen auch an den selben Strängen: Wenn wir Bedeutung betonen wollen, wählen wir ein Erstglied aus dessen Flexionsparadigma, meist ein genitivisches oder plurales. Hier ergänzen sich Bedeutungs- und Paradigmenregel.

Wie stellen wir Linguisten fest, welche Regel gerade greift? Im Einzelfall zu klären, welche Regel greift, wird vor allem dadurch erschwert, dass Zusammensetzungen von Individuen realisiert werden, die sich individuell zwischen den Regeln entscheiden. Wobei sie die Regeln „intuitiv und routiniert“ (Janich 2004, S. 110f.) anwenden, bei ausgeprägter Kompetenz auch „aufmerksam nutzen und reflektieren“ (ebd.).⁶² Woher wissen wir Linguisten, an welche Regel ein Individuum gerade denkt? Vielleicht, wenn wir in die individuellen Köpfe schauen könnten ...

Bis dahin schauen wir am besten ganz genau auf das, was uns fassbar ist. Wie schaut es denn zum Beispiel aus mit der Variation? Zunächst rein statistisch: Welche Rolle spielt Variation im Schnittstellenbereich von Zusammensetzungen überhaupt?

⁶² Auch diese Regeln „sind ohne Unterschied kulturell (und nicht etwa naturgesetzlich, wie dies – zumindest für den Bereich der grammatischen Regeln – teilweise unterstellt wird), das heißt, sie sind aktiv oder als (unbeabsichtigte) Handlungsnebenwirkungen im Rahmen von Sprachwandelprozessen von Menschen gemacht oder beeinflusst, um das Leben in einer kulturellen Gemeinschaft zu ermöglichen. Das impliziert, dass sie einerseits begründungspflichtig und revidierbar sind, erklärt aber auch, warum sie notwendigerweise eine so hohe Stabilität und Bindungskraft besitzen, wie es sich im täglichen Sprachhandeln und in Sprachbewertungen zeigt“ (Janich 2004, S. 55).

Augst (1975, S. 134) hat in seinem Belegmaterial 9,3% Erstglieder, die in zwei Varianten vorkommen, und 0,7% Erstglieder, die in drei Varianten vorkommen.⁶³ Nur vereinzelt, nämlich 8 mal, hat er in seinem Material Erstglieder, die in vier Varianten vorkommen.⁶⁴ Augsts Belegmaterial „erfasst 4025 verschiedene Substantive als Bestimmungswort in Zusammensetzungen“ (ebd.). Bei den DeReKo-Schnittstellenanalysen stimmen die Prozente der Erstglieder, die in zwei Varianten vorkommen wie *Arzt* und *Ärzte*, mit seinen überein, aber der Anteil der Erstglieder, die in drei oder vier Varianten vorkommen, ist prozentual deutlich höher. Analysiert hat Noah Bubenhofer verschiedene Erstglieder, die jeweils mehr als 10 mal im Korpus vorkommen.

Tabelle 11 zeigt, das wir es zu knapp 18,9% mit varianten Erstgliedern zu tun haben.⁶⁵ Davon variieren in den DeReKo-Schnittstellenanalysen 2% bei gleichem Zweitglied. Zum Beispiel hat das Erstglied *Blatt* in der Zusammensetzung mit dem Zweitglied *Menge* zwei Varianten: *Blattmenge* und *Blättermenge*. Noah Bubenhofer hat wiederum nur Erstglieder berücksichtigt, die im Korpus mehr als 10 mal vorkommen.

⁶³ Ich sage bewusst, dass ein Erstglied in zwei Varianten vorkommt, und nicht, dass ein Erstglied eine Variante (und folglich eine Normform) hat. Das Erstglied *Arzt* hat demnach nicht die Variante *Ärzte*, sondern kommt in den beiden Varianten *Arzt* und *Ärzte* vor. Denn auch für mich ist Standardsprache „eine genormte und kodifizierte Varietät“ (Spiekermann 2008, S. 27).

⁶⁴ Nach Augst (1975, S. 135) „bringt Grebe ein Wort mit 5 FMM: *Mannloch*, *Mannsbild*, *Manneswort*, *Mannentreue*, *Männertreu*. Hier liegen m.E. zwei polyseme Wörter *Mann-es*, U[mlaut]-*er* und *Mann-es*, *-en* vor.“ Was übrigens nicht nur ein Problem handverzettelter Linguisten ist, sondern auch ein Problem automatischer Analyse, denn der Rechner weiß nicht, dass die Erstglieder *Granat* in *Granatapfel* und *Granaten* in *Granatensalve* nur weitläufig miteinander zu tun haben, dass ein Mühlestein normalerweise nicht als Mühlenstein taugt vice versa und ein Bandchef nicht unbedingt ein Bandenchef sein muss vice versa. Auch analysiert er blindlings *Feuersturm* (*Feuer-sturm*) als *s*-Variante zu *Feuerturm* und verbucht *Beileidserklärung* unter *eid +s* oder *Vierfachaustausch* als Variante zu *Fächertausch* oder *Eiermarkt* als Variante zu *Eismarkt*.

Verschiedene Erstglieder	20 895	100%
Erstglieder, die in zwei Varianten vorkommen <i>Arzt</i> <i>Ärzte</i>	2 004	9,5%
Erstglieder, die in drei Varianten vorkommen <i>Betrieb</i> <i>Betriebe</i> <i>Betriebs</i>	1 706	8,1%
Erstglieder, die in vier Varianten vorkommen <i>Kind</i> <i>Kinds</i> <i>Kindes</i> <i>Kinder</i>	247	1,1%

Tabelle 11

⁶⁵ Erstglieder mit noch mehr Varianten fallen insgesamt unter die 0,5%-Grenze und sind meist auf Rechnerfehler zurückzuführen. So werden für das Erstglied *Bett* als Varianten-

bett + 0

bett + s

bett + en

bett + agssonn

bett + agssams

bett + agsonn

bett + agsams

Hier geht es um Belege wie *Doch in dem Schreiben, das die Hirten auf den gestrigen **Bettags**sonntag an ihre Schäflein schickten, bleibt es bei Andeutungen, bei allgemein gehaltenen Appellen* (Zürcher Tagesanzeiger 1999, IDS-Korpora). Der Lieblingsanalysefehler meiner Kollegin Sabina Ulsamer, die auch mit Rechneranalysen kämpft, ist übrigens die automatische Zerlegung von *Westeuropa* in *Weste* und *Uropa*.

Schnittstellenanalysen mit verschiedenem Erst- und Zweitglied	479 814	97,8%
Schnittstellenanalysen mit gleichem Zweitglied- und einem Erstglied in zwei Varianten <i>Blättermenge</i> <i>Blattmenge</i>	10 204	2,0%

Tabelle 12

Tabelle 12 ist zu ergänzen um die 0,1% Schnittstellenanalysen mit gleichem Zweitglied, deren Erstglied in drei Varianten vorkommt.⁶⁶ Der Rechner zählt dazu Zusammensetzungen wie *Taschendiebbande*, *Autodiebesbande*, *Autodiebsbande* mit den Varianten *Dieb*, *Diebes*, *Diebs*. Er hat *Dreieichpark*, *Eichepark*, *Eichenpark* sowie *Ämtergebäude*, *Amtsgebäude* und *Amtgebäude*. Letzteres ist überwiegend mit komplexem Erstglied belegt: *Forstamtgebäude*, *Postamtgebäude*,⁶⁷ kommt aber immerhin 7 mal als *Amtgebäude* vor, hauptsächlich in österreichischen Zeitungen.

*Auch das Argument mit dem (rund um das **Amtgebäude** nicht vorhandenen) „eigenen Wirkungsbereich“ zieht nicht.*

(Oberösterreichische Nachrichten 1997, IDS-Korpora).

Nübling/Szczepaniak (2009, S. 217) sehen in solchen gleichen Zusammensetzungen mit variierendem Erstglied typische Beispiele für „standardsprachliche Schwankungen“, die „echte Zweifelsfälle gemäß der Definition von Klein“ sind.⁶⁸ Sie nennen *Abiturfeier* und *Abitursfeier*, *Interessenausgleich* und *Interessensausgleich* sowie *Seminararbeit* und *Seminarsarbeit*.⁶⁹

⁶⁶ Darunter einige falsche Schnittstellenanalysen wie bei *Deutsch-Amerikanische-Institut*, *Deutsch-Amerikanischen-Institut* und *Deutsch-Amerikanisches-Institut* mit den nicht substantivischen Erstgliedern *Amerikanische*, *Amerikanischen* und *Amerikanisches*.

⁶⁷ Offenbar haben die Schreiber hier kein Bedürfnis gehabt, ihren Lesern gemäß der Strukturierungsregel eine Dekodierhilfe anzubieten.

⁶⁸ Nach Klein (2003, o.S.) ist ein „sprachlicher Zweifelsfall eine sprachliche Einheit (Wort/Wortform/Satz), bei der kompetente Sprecher im Blick auf (mindestens) zwei Varianten in Zweifel geraten können, welche der beiden Formen (standardsprachlich) korrekt ist (vgl. Sprachschwankung, Doppelform, Dublette). Die beiden Varianten eines Zweifelsfalls sind formseitig oft teildentisch“.

Bei den DeReKo-Schnittstellenanalysen lassen sich zahlreiche echte Zweifelsfälle ausmachen. Um nur die daran beteiligten Erstglieder mit dem Anfangsbuchstaben *A* zu zeigen – wohlgemerkt: Es sind nur die echten Zweifelsfälle des Typs *Abfahrtsmöglichkeit* neben *Abfahrtsmöglichkeit*!

<i>Abfahrt</i>	<i>Aktuar</i>	<i>Antrag</i>	<i>Aufwand</i>
<i>Abfall</i>	<i>Alp</i>	<i>Antrieb</i>	<i>Aufzucht</i>
<i>Abgabe</i>	<i>Altar</i>	<i>Antwort</i>	<i>Aufzug</i>
<i>Abgeltung</i>	<i>Alter</i>	<i>Anwalt</i>	<i>Auge</i>
<i>Abguss</i>	<i>Altertum</i>	<i>Anzeige</i>	<i>Auktion</i>
<i>Abitur</i>	<i>Amateur</i>	<i>Anzug</i>	<i>Ausbruch</i>
<i>Abonnement</i>	<i>Aminosäure</i>	<i>Aorta</i>	<i>Ausdruck</i>
<i>Absage</i>	<i>Amt</i>	<i>Apartheid</i>	<i>Ausfahrt</i>
<i>Abschied</i>	<i>Analyse</i>	<i>Apotheker</i>	<i>Ausfall</i>
<i>Abschlag</i>	<i>Andrang</i>	<i>Aquarium</i>	<i>Ausflug</i>
<i>Abschluss</i>	<i>Anfahrt</i>	<i>Arbeit</i>	<i>Ausgabe</i>
<i>Abschnitt</i>	<i>Anfall</i>	<i>Armut</i>	<i>Ausgang</i>
<i>Abschwung</i>	<i>Anfang</i>	<i>Art</i>	<i>Ausgleich</i>
<i>Absicht</i>	<i>Anfrage</i>	<i>Arthrose</i>	<i>Aushang</i>
<i>Abstand</i>	<i>Angabe</i>	<i>Arzt</i>	<i>Auskunft</i>
<i>Abstieg</i>	<i>Angebot</i>	<i>Asche</i>	<i>Ausland</i>
<i>Abzug</i>	<i>Angriff</i>	<i>Attentat</i>	<i>Ausnahme</i>
<i>Achse</i>	<i>Anhang</i>	<i>Aufbruch</i>	<i>Ausschnitt</i>
<i>Acker</i>	<i>Ankauf</i>	<i>Aufenthalt</i>	<i>Ausschuss</i>
<i>Adel</i>	<i>Ankunft</i>	<i>Auffahrt</i>	<i>Aussicht</i>
<i>Ader</i>	<i>Anlage</i>	<i>Aufgabe</i>	<i>Ausstieg</i>
<i>Adler</i>	<i>Anlauf</i>	<i>Auflage</i>	<i>Austausch</i>
<i>Admiral</i>	<i>Anlauf</i>	<i>Aufnahme</i>	<i>Austrag</i>
<i>Admiralität</i>	<i>Anleihe</i>	<i>Aufschlag</i>	<i>Austritt</i>
<i>Advent</i>	<i>Anschluss</i>	<i>Aufschwung</i>	<i>Auszug</i>
<i>Aggregat</i>	<i>Anspruch</i>	<i>Aufsicht</i>	<i>Automobil</i>
<i>Akt</i>	<i>Anstalt</i>	<i>Aufstieg</i>	<i>Autor</i>
<i>Aktion</i>	<i>Anteil</i>	<i>Auftakt</i>	
<i>Aktivität</i>	<i>Antike</i>	<i>Auftrag</i>	

⁶⁹ Echte Zweifelsfälle sind triftige Gründe gegen die Nutzung allgemeiner Wörterbücher zur Erforschung von Varianten. Ich kenne jedenfalls kein Wörterbuch, das alle Varianten angibt und statistisch verortet. Immerhin findet man in Variantenwörterbüchern wie Ammon et al. (2004) den einen oder anderen Hinweis auf Schnittstellenvarianz, etwa regionales *Fabrikarbeiter* neben regionalem *Fabrikarbeiter*.

Wie variieren wir Erstglieder wie diese denn nun? Welche Wortformen und Fugenelemente konkurrieren?

Zur besseren Übersicht habe ich alle Erstglieder, die an echten Zweifelsfällen beteiligt sind, nach Art der Varianz sortiert und jeweils vier typische Belege und deren Vorkommenshäufigkeit recherchiert. Die dominierenden Formen habe ich fett gesetzt. Gerechnet habe ich in Tabelle 13 wiederum mit Tokens, weil es bei Zusammensetzungen mit gleichem Zweitglied keinen Sinn macht, mit Types zu rechnen: Ich würde ja zum Beispiel genau ein Type *Ankaufpreis* und ein Type *Ankaufspreis* erhalten und wüsste nichts über Dominanzen.

0 neben s	<i>Ankaufpreis</i>	30
	<i>Ankaufspreis</i>	148
	<i>Mittagessen</i>	37 322
	<i>Mittagessen</i>	348
	<i>Umbruchphase</i>	682
	<i>Umbruchsphase</i>	96
	<i>Zugunglück</i>	3 349
	<i>Zugsunglück</i>	412
0 neben n	<i>Anleihehändler</i>	38
	<i>Anleihenhändler</i>	15
	<i>Antikerezeption</i>	15
	<i>Antikenrezeption</i>	50
	<i>Messechef</i>	1 335
	<i>Messenchef</i>	86
	<i>Speisekarte</i>	10 273
	<i>Speisenkarte</i>	418
0 neben e	<i>Arzthaus</i>	168
	<i>Ärztehaus</i>	2 330
	<i>Fruchtbowle</i>	18
	<i>Früchtebowle</i>	18
	<i>Kraftfeld</i>	799
	<i>Kräftefeld</i>	155
	<i>Mausfamilie</i>	16
	<i>Mäusefamilie</i>	74

0 neben en	Bankgeheimnis	6 805
	<i>Bankengeheimnis</i>	42
	Miniaturmaler	39
	<i>Miniaturenmalers</i>	19
	Straftatserie	49
	<i>Straftatenserie</i>	37
	<i>Zinssaldo</i>	68
	Zinsensaldo	69
0 neben er	Bildauswahl	343
	<i>Bilderauswahl</i>	65
	Buchshop	64
	<i>Büchershop</i>	26
	<i>Kornkasten</i>	50
	Körnerkasten	74
	Wortverzeichnis	337
	<i>Wörterverzeichnis</i>	112
0 neben es	<i>Geistmenschen</i>	25
	Geistesmenschen	194
	<i>Gesetzinitiative</i>	17
	Gesetzesinitiative	2 822
	Heerlager	495
	<i>Heereslager</i>	17
	Meerenge	1 390
	<i>Meeresenge</i>	134
0 neben e-Tilgung plus s	Aufnahmeprüfung	3 354
	<i>Aufnahmsprüfung</i>	431
	Gewebeschicht	21
	<i>Gewebsschicht</i>	15
	<i>Hilfeappell</i>	11
	Hilfsappell	281
	Pferdekopf	381
	<i>Pferdskopf</i>	23

s neben en	<i>Fakultätstage</i>	27
	<i>Fakultätentage</i>	133
	<i>Liegenschafts- bestand</i>	17
	<i>Liegenschaften- bestand</i>	11
		33
		14
	<i>Optionsmarkt</i>	11
	<i>Optionenmarkt</i>	14
s neben es	<i>Variationskette</i>	
	<i>Variationenkette</i>	
	<i>Diebsbeute</i>	265
	<i>Diebesbeute</i>	320
	<i>Kindsmörder</i>	174
	<i>Kindesmörder</i>	126
	<i>Kindsmutter</i>	233
	<i>Kindesmutter</i>	514
s neben er	<i>Landsleute</i>	26 521
	<i>Landesleute</i>	41
	<i>Amtsgang</i>	20
	<i>Ämtergang</i>	12
	<i>Mitgliedsliste</i>	53
	<i>Mitgliederliste</i>	560
	<i>Rindsgeschnetztes</i>	21
	<i>Rindergeschnetztes</i>	12
s neben e	<i>Volksgemeinschaft</i>	1 123
	<i>Völkergemeinschaft</i>	1 770
	<i>Bestandsübersicht</i>	6
	<i>Beständeübersicht</i>	1
	<i>Hundskopf</i> ⁷⁰	48
	<i>Hundekopf</i>	102
	<i>Schweinsohr</i>	35
	<i>Schweineohr</i>	40
	<i>Tagsablauf</i>	2
	<i>Tageablauf</i> ⁷¹	3

e-Tilgung neben n	Erdball	4 080
	Erdenball	32
	Farbauswahl	265
	Farbenauswahl	11
	Kirchturm	10 409
	Kirchenturm	148
	Palmbaum	78
	Palmenbaum	35

Tabelle 13

Auch hier zeigt sich der Vorteil großer Korpora: Wir Linguisten können uns einen Überblick verschaffen. Zunächst habe ich mir in Tabelle 13 einen Überblick darüber verschafft, welche Wortformen beziehungsweise Fugenelemente überhaupt untereinander variieren. Dabei bestätigen die DeReKo-Schnittstellenanalysen die Idee von Michel (2010, S. 17): Am häufigsten sind die Null- und die *s*-Form; entsprechend können wir deren häufiges Variieren miteinander und mit anderen Formen erwarten.

Die Nullform, die über 70% der Erstglieder ausmacht, variiert tatsächlich mit allen anderen Formen.⁷² Keine andere Form hat so viele verschiedene konkurrierende Varianten.

	<i>s</i>	<i>n</i>	<i>es</i>	<i>en</i>	<i>er</i>	<i>e</i>
0	+	+	+	+	+	+

Tabelle 14

Aber auch die *s*-Form konkurriert mit fast allen anderen Formen.

⁷⁰ Darunter zahlreiche Belege für *Hundskopf* als Name des Berges im Karwendel, aber auch gleichbedeutend mit *Hundekopf*. Jetzt steht das orange mit aufgemalten Mauerresten, einer Echse und einem **Hundskopf** verzierte und mit Graffiti-Schriftzug- und -Muster besprayed Häuschen auf dem Pausenplatz (St. Galler Tagblatt 2001, IDS-Korpora).

⁷¹ Zum Beispiel: *Der Obstbau, die Bienenzucht und „werkeln“ in seiner Werkstatt füllen den Tagsablauf des über 90jährigen Heinrich aus* (Vorarlberger Nachrichten 1998, IDS-Korpora) und *Das deutsche Vormittagsschulmodell hat in Frankreich gegenüber früher an Attraktivität verloren. Die im Tageablauf praktizierte Extensivität findet sich aber im Wochen- und Jahresrhythmus nicht wieder* (Frankfurter Allgemeine 2003, IDS-Korpora).

⁷² Die zum Teil komplizierten Tilgungsformen wie die *e*-Tilgung plus *s* in *Aufnahmsprüfung* vernachlässige ich hier der Übersicht wegen.

	0	<i>es</i>	<i>en</i>	<i>er</i>	<i>e</i>
<i>s</i>	+	+	+	+	+

Tabelle 15

Nicht variiert wird *s* mit *n*. Das lässt sich unter anderem flexionsparadigmisch erklären: Die *n*-Form ist typischerweise eine Variante zu maskulinen Erstgliedern auf *e* wie *Affe* neben *Affen*, zu femininen Erstgliedern auf *e* wie *Achse* neben *Achsen*. Die maskulinen und femininen Erstglieder haben kein *s* in ihrem Paradigma. Außerdem kommt die *n*-Form paradigmisch vor bei Erstgliedern auf *er* wie *Bauer* neben *Bauern* und *Oper* neben *Opern* sowie bei Erstgliedern auf *el* wie *Achsel* neben *Achseln* und *Fackel* neben *Fackeln*, die ebenfalls keine *s*-Form in ihrem Paradigma haben. Nur die neutralen Erstglieder auf *e* wie *Auge* neben *Augen* haben zwar neben dem *n* auch ein *s* in ihrem Paradigma, in Zusammensetzungen kommt die *s*-Form dennoch nicht als Variante vor.

Alle weniger häufigen Erstgliedformen variieren immer mit der Null- oder der *s*-Form, selbst wenn ihr Paradigma auch andere Varianten zulassen würde. Auch bei Erstgliedern mit mehr als zwei Varianten ist immer eine Null- oder *s*-Form oder beides dabei, zum Beispiel bei *Tag*, *Tages*, *Tage*, bei *Betrieb*, *Betriebe*, *Betriebs*, bei *Kind*, *Kinds*, *Kindes*, *Kinder*.

	0	<i>s</i>
<i>n</i>	+	
<i>es</i>	+	+
<i>en</i>	+	+
<i>er</i>	+	+
<i>e</i>	+	+

Tabelle 16

Intuitiv daraus geschlossen hat Michel (2010, S. 17), dass Erstglieder in weniger häufigen und gebräuchlichen Formen „nicht variieren (*Kind* + *es* + *missbrauch* wird innerhalb der Standardsprache kaum mit *Kind* + *er* + *missbrauch* variieren)“. Schauen wir uns zunächst sein Beispiel an. In den IDS-Korpora der geschriebenen Sprache findet sich *Kindermisbrauch* 48 mal.

*Wen überfällt nicht Ekel und Empörung angesichts von **Kindermisbrauch**?*

(St. Galler Tagblatt 2008, IDS-Korpora)

***Kindermisbrauch** durch Priester sei ein Verbrechen, das auch die Staatsanwaltschaft verfolgen müsse.*

(Die Zeit 2005, IDS-Korpora)

Das ist als Variante zu 1 726 mal *Kindesmissbrauch* wenig, aber die Variante ist da. Im September 2010 habe ich *Kindermisbrauch* sogar rund 22 900 mal gegoogelt neben rund 265 000 mal *Kindesmissbrauch*. Offenbar sollten wir mit viel mehr Varianz rechnen. Einer Varianz, die Lerner freuen sollte: Wo viel möglich und gebräuchlich ist, kann man weniger verkehrt machen.⁷³

Wie sieht es mit Michels intuitivem Gedanken statistisch aus? Schauen wir uns zunächst die Nullform an. Sie nimmt in den DeReKo-Schnittstellenanalysen mit über 70% die Spitze des Rankings ein, wie es sich in Tabelle 17 aus Tabelle 2 ergibt.

Ranking der Schnittstellen
1. Platz: 0
2. Platz: s
3. Platz: n
4. Platz: es
5. Platz: en
6. Platz: er
7. Platz: e

Tabelle 17

Variiert die Nullform auch statistisch besonders häufig mit der ebenfalls häufigen s-Form, die mit über 16% den zweiten Rankingplatz einnimmt? Und wie häufig variiert sie mit weniger häufigen Formen? Gerechnet hat Noah Bubenhofner alle Nullformen mit mindestens einer Variante. Die Varianten habe ich nach ihrem Platz in der Rankingliste angeordnet.

⁷³ Diese Rechnung mache ich ohne den Lerner, wenn stimmt, was Figueras (2010, S. 203) sagt: „Die Leute [...] schätzen es auch nicht, wenn ihnen grenzenlose Kreativität ermöglicht wird, die bei ihnen Beklemmungen hervorruft. Sie ziehen knappe, klare Anweisungen vor, die ihnen keinen großen Handlungsspielraum lassen.“

Nach Diagramm 1 variiert tatsächlich die Nullform mit der *s*-Form am häufigsten. So weit trägt Michels Intuition nicht. Bei den weniger häufigen Formen allerdings verschiebt sich das Ranking: So hat zum Beispiel die *n*-Form den Rankingplatz 3, variiert mit der Nullform aber fast genauso häufig wie die Platz 4 belegende *es*-Form, jedenfalls wenn man die Tokens zählt. Zählt man die Types, variiert die *es*-Form mit der Nullform weniger als die auf den hinteren Plätzen rangierenden *er*- und *e*-Formen.

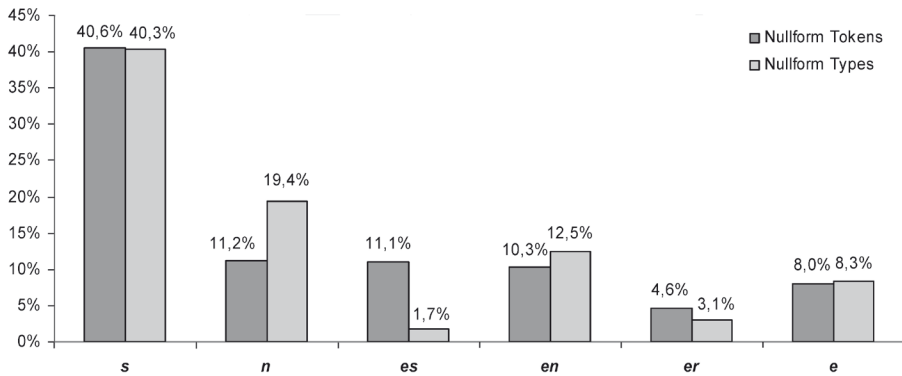


Diagramm 1

Dieser Unterschied, den mitunter das Rechnen mit Tokens und Types ergibt, ist natürlich auffällig, besonders der krasse Unterschied bei der Variation von Null- mit *es*-Form. Wird die Statistik verzerrt von einzelnen Exemplaren der *es*-Form?⁷⁴ Zum Beispiel hat das Erstglied *Bund* in seiner *es*-Form *Bundes* einen extrem hohen Anteil an der Varianz von *es*-Form neben Nullform. Hier geht es um 93 330 Tokens für Zusammensetzungen wie *Bundespräsident* neben 68 Tokens für Zusammensetzungen wie *Sportbundpräsident*.⁷⁵ Als Tokens gezählt ergibt sich also ein deutliches Übergewicht von *Bundes*; als Types gezählt zählt in der Gesamtstatistik genau ein Type *Bundespräsident* neben genau einem Type *Sportbundpräsident*. Sollte man also einzelne, besonders häufig vorkommende Schnittstellen ausblenden, indem man nur

⁷⁴ „Verzerrt ist das Gegenteil von repräsentativ: Gewisse Teilmengen der Grundgesamtheit [...] sind in der Stichprobe systematisch überrepräsentiert, andere sind systematisch unterrepräsentiert.“ (Krämer 2000, S. 98).

⁷⁵ Dabei lasse ich *Bund* im juristischen Sinne von ‘Verband als Gruppe von Einzelpersonen oder Körperschaften aller Art’ gelten. Fehlerhaft hat der Rechner allerdings *BUND* als Kurzwort für *Bund für Umwelt und Naturschutz Deutschland e.V.* miteinbezogen. Statistisch macht sich das zum Beispiel bemerkbar bei 134 340 Belegen für *Bundesrat* neben 14 Belegen für *BUND-Ratgeber*.

Types zählt? Noah Bubenhofer hat dazu angemerkt, dass ja aber einzelne, besonders häufig vorkommende Schnittstellen auch unsere Schnittstellengestaltung besonders beeinflussen. Dann wäre die Zahl der Tokens ebendoch aussagekräftiger.

Schauen wir uns nun noch die *s*-Form an. Mit welchen Formen variiert sie am häufigsten? Die Varianten habe ich wiederum nach ihrem Platz auf der Rankingliste angeordnet. Wiederum sinken die Anteile zwischen Tokens und Types besonders bei der *es*-Form. Die von Michel vermutete Tendenz, dass die häufigsten Formen am häufigsten mit anderen häufigsten Formen variieren, vor allem die Null- mit der *s*-Form, zeigt sich auch in Diagramm 2.

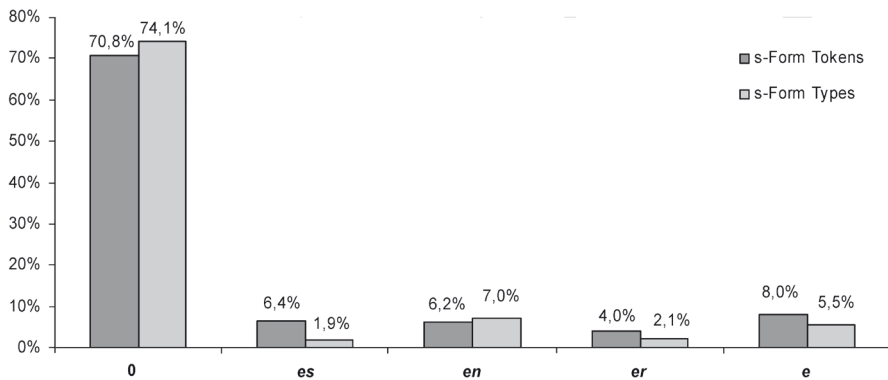


Diagramm 2

Die häufige Konkurrenz der *s*-Form mit der Nullform kann man auch daran erkennen, was Laien vom Institut für Deutsche Sprache wissen wollen. Hier drei typische Anfragen.

Mail vom 7.5.2007

Sehr geehrte Frau Dr. Trabold,

ich benötige Ihre Hilfe in bezug auf die Verwendung des sog. „Fugen-S“.

Es herrscht Uneinigkeit über die Verwendung des Fugen-S bei dem Wort „Rück-
kauf(s)preis“.

Aus meiner Sicht wird bei diesem Wort das „Fugen-S“ verwendet.

Mit freundlichen Grüßen

M.H.

Leiter Freiwilliger Umtausch

Mail vom 22.10.2007

Sehr geehrte Frau Trabold,
ich bitte Sie um kurze Hilfe. Bekommt das Wort „Werkbesichtigung“ ein Fugen s? Also „Werksbesichtigung“, in meinen Lexika konnte ich dies nicht finden.

Vielen, vielen Dank!

Mit freundlichen Grüßen, K.N.
Unternehmenskommunikation

Mail vom 1.12.2010

Sehr geehrte Frau Dr. Trabold,
bevor sich unsere grammatikalische Diskussion in einen herzhaften Streit verwandeln, habe ich beschlossen, mich an das Institut für Deutsche Sprache zu wenden.

Vielleicht können Sie ja mit einer kurzem Antwort, die entscheidende Klarheit herstellen.

Unsere Frage ist recht einfach:

Schreibt man korrekt „Betriebsratscout“ - also der Scout für den Betriebsrat?

Oder heißt es „Betriebsratsscout“ - also der Scout des Betriebsrats?

Ich würde mich über eine Antwort von Ihnen sehr freuen. Außerdem bin ich gespannt, wer von uns Recht hat.

Mit freundlichen Grüßen

A.G.

Mit Varianz des Erstglieds ist jedenfalls nur in den geschilderten Fällen zu rechnen. Auch das ist eine gute Nachricht: Es reduziert die Möglichkeiten von Varianz.

Kann der Rechner außerdem noch rausrechnen, welche Varianten genau dominieren? Zunächst fällt auf, dass nur in der letzten Zelle der Tabelle 13 eine Variante durchgehend dominiert: Die Tilgung des Schwa dominiert in allen vier Belegen über die paradigmatische Form mit *n*: *Erd*, *Farb*, *Kirch*, *Palm* dominiert über *Erden*, *Farben*, *Kirchen* und *Palmen*. In allen anderen Zellen der Tabelle 13 ist mal die eine, mal die andere Variante dominant.

Hält sich die Dominanz der schon historischen Schwa-Tilgung *Erde* zu *Erd* auch bei anderen Erstgliedern durchgängig? Sie zeigt sich auch in meinen weiteren Recherchen, diesmal nach dem Erstglied insgesamt.⁷⁶

Achs	2 619
<i>Achsen</i>	524
Eck	37 803
<i>Ecken</i>	2 688
Eich ⁷⁷	23 078
<i>Eichen</i>	6 961
Filial	17 548
<i>Filialen</i>	17
Grenz	134 211
<i>Grenzen</i>	108
Kapell	5 425
<i>Kapellen</i>	4 585
Lokomotiv	3 891
<i>Lokomotiven</i>	127
Münz	11 728
<i>Münzen</i>	608
Perspektiv	1 643
<i>Perspektiven</i>	741
Spiral	863
<i>Spiralen</i>	12
Sprach	83 280
<i>Sprachen</i>	4 085

Tabelle 18

Tabelle 18 zeigt, dass wir im Deutschen dazu tendieren, das Schwa femininer Erstglieder zu tilgen. Ich werde nun aber nicht empfehlen, das Schwa femininer Erstglieder immer zu tilgen, weil man damit immer auf der Seite des mehrheitlichen Sprachgebrauchs steht. Die Schwa-Tilgung ist eine Tendenz,

⁷⁶ Nicht berücksichtigt habe ich hier mehrdeutige Erstglieder. So kann bei *Erb* versus *Erben* das *Erb* nicht nur als substantivisches *Erbe* mit *e*-Tilgung, sondern auch als verbales Erstglied bestimmt werden. So auch die Erstglieder *Lehr*, *Leucht*, *Miet* oder *Straf*.

⁷⁷ Manche Erstglieder sind besonders häufig in Eigennamen eingebunden. So hat der Mannheimer Morgen 1 534 mal *Eichbaum*, die Frankfurter Allgemeine dagegen nur 8 mal. Der Unterschied erklärt sich daraus, dass es in Mannheim eine Brauerei *Eichbaum* gibt.

keine Regel. Dass es nämlich Gründe gibt, nicht jedes Schwa femininer Erstglieder zu tilgen, ist leicht nachvollziehbar: So kann man zum Beispiel mit *Sprachenkenntnis* anderes ausdrücken als mit *Sprachkenntnis*.

Schauen wir uns noch eine andere Zelle genauer an: Die Nullformen in *Mittagessen* und *Zugunglück* dominieren über die *s*-Formen. Bei *Ankaufpreis* und *Ankaufspreis* dominiert aber genau umgekehrt die *s*-Form. Zufall? Auch in Zusammensetzungen mit dem Erstglied *Abfahrt* dominiert mehr oder weniger deutlich die *s*-Form. Zur Sicherheit habe ich im November 2010 auch noch einmal nachgegoogelt. Google hat ebenfalls eine klare *s*-Dominanz.

	DeReKo	Google
<i>Abfahrtmöglichkeit</i>	14	2 630
<i>Abfahrtsmöglichkeit</i>	35	22 400
<i>Abfahrttrampe</i>	14	2 830
<i>Abfahrtsrampe</i>	195	20 200

Tabelle 19

Liegt die Tendenz in Tabelle 19 daran, dass die Erstglieder *Ankauf* und *Abfahrt* abgeleitete Substantive sind? Sie sind Ableitungen aus Präverbfügungen.⁷⁸ Präverbfügungen bestehen aus einem nur lose verbundenem Präverb wie *an* und einem Verb wie *kaufen* und haben alle einen besonderen Akzent: *ánkaufen*, *ábfahren*. Auch andere aus Präverbfügungen abgeleitete Erstglieder wie *Anfahrt*, *Angriff*, *Anschlag* zeigen sowohl bei den DeReKo-Schnittstellenanalysen als auch bei Google *s*-Dominanz.⁷⁹

	DeReKo	Google
<i>Anfahrtweg</i>	30	4 900
<i>Anfahrtsweg</i>	1 120	175 000
<i>Angriffkrieg</i>	14	1 370
<i>Angriffskrieg</i>	2 151	462 000
<i>Anschlaggefahr</i>	48	1 120
<i>Anschlagsgefahr</i>	121	39 300

Tabelle 20

⁷⁸ Auch Partikelverben genannt. Vgl. Donalies (1999).

⁷⁹ Und auch zu Ableitungen aus Präfixverben wie *Bericht*, deren Akzentverhältnisse allerdings von den Präverbfügungen abweichen: *berichten*. In den DeReKo-Schnittstellenanalysen stehen 13 Belege *Berichtjahr* deutlich abgehoben 3 262 Belegen *Berichtsjahr* gegenüber.

Dann müsste *Umbruch* die gleichen Zahlen liefern. Aber bei *Umbruch* in Tabelle 21 dominiert en detail die Nullform, genauso wie bei *Mittag* und *Zug*. Übrigens unabhängig davon, ob das Zweitglied mit einem *s* anlautet oder nicht.

	DeReKo	Google
<i>Umbruchjahr</i>	43	9 910
<i>Umbruchsjaar</i>	22	7 000
<i>Umbruchprozess</i>	81	9 830
<i>Umbruchsprozess</i>	14	5 830
<i>Umbruchsituation</i>	448	19 500
<i>Umbruchssituation</i>	44	8 430
<i>Umbruchstimmung</i>	90	11 500
<i>Umbruchsstimmung</i>	13	3 410

Tabelle 21

Behandeln wir also Erstglieder, die aus Präverbfügungen mit *um* abgeleitet sind, anders als solche, die aus Präverbfügungen mit *ab* oder *an* abgeleitet sind? *Umschlag* sieht in Tabelle 22 zunächst nach einer Bestätigung dieser Vermutung aus.

	DeReKo
<i>Umschlaggesellschaft</i>	14
<i>Umschlaggesellschaft</i>	35

Tabelle 22

Aber schon Google widerspricht.⁸⁰

	Google
<i>Umschlaggesellschaft</i>	18 500
<i>Umschlaggesellschaft</i>	9 130

Tabelle 23

Und auch in den DeReKo-Schnittstellenanalysen finden sich Nullformdominanz bei Erstglied *Umschlag*.

⁸⁰ Einen Grund für die Divergenz kann ich nicht sehen: Die Zusammensetzungen kommen in den DeReKo-Schnittstellenanalysen vor allem in Zeitungen vor. Die Zeitungen sind regional und stilistisch breit gestreut. Sie entsprechen der Korpuslage von *Umschlaggesellschaft*.

	DeReKo	Google
<i>Umschlaghalle</i>	53	15 700
<i>Umschlagshalle</i>	26	7 080

Tabelle 24

Die Tabellen 19 bis 24 demonstrieren also, dass es bei variierenden Erstgliedern sowohl konstante als auch wechselnde Dominanzen gibt. Erstglieder eines bestimmten Typs tendieren mal konstant zu einer Variante, mal wechseln die Dominanzen bei solchen Erstgliedern, mal wechseln sie bei ein und dem selben Erstglied. Das sieht nach regellosem Chaos aus.⁸¹ Oder lässt sich mit Hilfe der Regeln doch noch halbwegs Geordneteres rausrechnen?

Die sich widerstreitenden, sich ergänzenden Regeln führen zu Varianten. Schauen wir uns also die Varianten unter den Regelaspekten genauer an und zwar die

- paradigmisch bedingten Varianten,
- bedeutungsbedingten Varianten,
- euphoniebedingten Varianten,
- strukturierungsbedingten Varianten,
- gewohnheitsbedingten Varianten.

Paradigmisch bedingte Varianten

Die Erstglieder in Zusammensetzungen entsprechen entweder ihrem Flexionsparadigma oder nicht. Die einzige unparadigmische Form ist die unparadigmische *s*-Form.

Bei den DeReKo-Schnittstellenanalysen variieren mit dem unparadigmischen *s* nur zwei paradigmische Formen: Überwiegend ist es die Nullform,⁸² mitunter aber auch die plurale Form mit *en*. Nebeneinander als Erstglieder gebräuchlich sind zum Beispiel *Abfahrt* neben *Abfahrts* und *Aktionen* neben *Aktions*.

⁸¹ „Sie mögen mir entgegenhalten, dass diese krasse Statistik nichts allzu Erfreuliches enthält. Vielleicht, aber so ist das Leben, und ich warne Sie lieber gleich, dass diese Abhandlung mit dem Leben nicht viel Federlesens macht“ (Vargas 2009, S. 12).

⁸² Bei den Varianten Nullform und Fugen-*s*-Form bin ich mir nicht sicher, ob es sich nicht einfach um Verschreibungen handelt, etwa bei 18 Belegen für *Arbeitswelt* gegenüber 16 093 Belegen für *Arbeitswelt*. Besonders bei Zweitgliedern mit einem initialen *s* kommen solche Verschreibungen vermutlich häufiger vor, so bei den 29 Belegen für *Sicherheitsstandard* gegenüber 6 435 Belegen für *Sicherheitsstandard*.

0 neben unparadigmischem <i>s</i>	<i>Abfahrt</i>	439
	<i>Abfahrts</i>	11 392
	<i>Aufsicht</i>	696
	<i>Aufsichts</i>	172 192
	<i>Auskunft</i>	299
	<i>Auskunfts</i>	4 655
	<i>Fraktion</i>	504
	<i>Fraktions</i>	252 385
en neben unparadigmischem <i>s</i>	<i>Aktionen</i>	24
	<i>Aktions</i>	66 285
	<i>Antiquitäten</i>	3 982
	<i>Antiquitäts</i>	1
	<i>Konföderationen</i>	335
	<i>Konföderations</i>	148
	<i>Minderheiten</i>	7 095
	<i>Minderheits</i>	15 728

Tabelle 25

Bei den von mir recherchierten Null- neben unparadigmischen *s*-Formen der Tabelle 25 dominieren durchgehend die *s*-Formen.⁸³ Bei den *en*-Formen neben unparadigmischen *s*-Formen dominieren mal die einen, mal die anderen. Dabei schwanken die Dominanzen selbst bei gleich endenden Erstgliedern. Als Beispiel habe ich in Tabelle 26 Erstglieder auf *ion* recherchiert.

Tabelle 26 sieht nach einer klaren Dominanz der unparadigmischen *s*-Fuge aus, die im Übrigen auch meine Einzelrecherchen ergibt. Selbst bei Erstgliedern mit allgemeiner *en*-Dominanz dominiert im Einzelfall oft die *s*-Form, so beim Erstglied *Generation* in den Zusammensetzungen *Generationenwechsel* und *Generationswechsel*, die einen großen Anteil an den Zusammensetzungen mit *Generation* ausmachen. So auch bei den weniger häufigen Zusammensetzungen *Generationenerfahrung* und *Generationserfahrung* (siehe Tab. 27).

⁸³ Bei den von mir recherchierten Nullformen neben paradigmischen Genitiv-*s*-Formen dominieren übrigens mitunter auch die Nullformen: So stehen zum Beispiel 14 432 mal *Anteil* neben 11 085 mal *Anteils* oder 4 374 mal *Apartheid* neben 523 *Apartheids*. Auch in den einzelnen Zusammensetzungen dominiert die Nullform: 1 299 mal ist *Apartheidregime* belegt neben 138 mal *Apartheidregime*.

Dominanz en vor s		Dominanz s vor en	
Generationen	15 162	Aktionen	24
<i>Generations</i>	8 364	Aktions	66 285
Konföderationen	335	<i>Kommissionen</i>	9
<i>Konföderations</i>	148	Kommissions	27 837
		<i>Optionen</i>	1 355
		Options	4 851
		<i>Regionen</i>	923
		Regions	6 409
		<i>Sanktionen</i>	369
		Sanktions	3 355
		<i>Stationen</i>	1 095
		Stations	5 177
		<i>Variationen</i>	365
		Variations	947
		<i>Visionen</i>	59
		Visions	1 708

Tabelle 26

<i>Generationenerfahrung</i>	13
Generationserfahrung	35
<i>Generationenwechsel</i>	2 458
Generationswechsel	6 185

Tabelle 27

Das heißt: Zwar greift in den meisten Fällen die Paradigmenregel, denn zu über 90% sind die Formen ja paradigmisch; die unparadigmische *s*-Fuge hat einen geringen Anteil von gut 8% aller DeReKo-Schnittstellenanalysen. Aber sobald die unparadigmische *s*-Fuge gesetzt wird, dominiert sie.

Bedeutungsbedingte Varianten

Bei den DeReKo-Schnittstellenanalysen sind 29,5 % der Erstglieder sichtbar durch ein paradigmatisches Flexions- oder ein unparadigmatisches Fugenelement markiert. Mitunter wird mit dieser Markierung Bedeutung angezeigt. Dabei geht es vor allem um Betonung von Vielzahl und Zugehörigkeit. Vielzahl wird

durch Pluralmarker, Zugehörigkeit durch Genitivmarker angezeigt.⁸⁴ Bedeutungsbedingt entsteht Varianz der Erstglieder demnach in drei Kombinationsfällen.

Nichtmarkierung neben Markierung als vielzählig	<i>Embryoforschung</i>	17
	<i>Embryonenforschung</i>	34
	<i>Speisekarte</i>	10 549
	<i>Speisenkarte</i>	432
Nichtmarkierung neben Markierung als zugehörig	<i>Farmerfamilie</i>	84
	<i>Farmersfamilie</i>	26
	<i>Schafffleisch</i>	302
	<i>Schaffsfleisch</i>	38
Markierung als vielzählig neben Markierung als zugehörig⁸⁵	<i>Direktorenposten</i>	209
	<i>Direktorsposten</i>	176
	<i>Schweineohr</i>	40
	<i>Schweinsohr</i>	35

Tabelle 28

Wie genau machen wir es mit der Markierung von Bedeutung? Wann wählen wir bedeutungsbedingt Varianten?

Linguisten beobachten mitunter „semantische Differenzierung“, unter anderen Ribeiro (1996, S. 15f.) bei *Kohle* neben *Kohlen*: Sie liest „Kohle als Rohstoff“ wie bei *Kohlepapier* versus Kohle als zählbares Material wie bei *Kohleneimer*. Zählbarkeit setzt Vielzahl voraus; bei zählbarem Material sind also plurale Erstglieder zu erwarten. Was sagen die DeReKo-Schnittstellenanalysen darüber aus? Dazu habe ich sämtliche echte Zweifelsfälle mit den Erstgliedvarianten *Kohle* neben *Kohlen* recherchiert. Sinnvollerweise habe ich auch hier wieder Tokens gezählt, denn Types gäbe es ja jeweils nur zwei, nämlich ein Type *Kohlepapier* und ein Type *Kohlenpapier*.

⁸⁴ Es gilt aber nicht der Umkehrschluss: Erstglieder mit Plural- und Genitivmarker zeigen nicht immer Vielzahl und Zugehörigkeit an.

⁸⁵ Pluralformen können zwar auch Zugehörigkeit markieren (*die Posten der Direktoren, die Ohren der Schweine*); ich gehe aber davon aus, dass bei Pluralformen in erster Linie Vielzahl betont werden soll.

Tendenz zu <i>Kohle</i>		Tendenz zu <i>Kohlen</i>		Gleichgewicht zwischen <i>Kohle und Kohlen</i>	
Kohlepapier	75	<i>Kohlehandlung</i>	17	Holzkohlegrill	338
<i>Kohlenpapier</i>	1	Kohlenhandlung	115	Holzkohlengrill	364
Kohleherd	54	<i>Kohlekeller</i>	22	Kohleschuppen	11
<i>Kohlenherd</i>	21	Kohlenkeller	275	Kohlenschuppen	13
Kohleabbau	947	<i>Kohleeimer</i>	3		
<i>Kohlenabbau</i>	49	Kohleneimer	16		
Kohlefaser	696				
<i>Kohlenfaser</i>	19				
Kohleverbrauch	114				
<i>Kohlenverbrauch</i>	16				
Kohleofen	356				
<i>Kohlenofen</i>	13				
Kohlekraftwerk	7 336				
<i>Kohlenkraftwerk</i>	74				

Tabelle 29

Die Tabelle 29 ergibt zahlreiche Belege für die Varianz der von Ribeiro genannten Zusammensetzungen. Die Zusammensetzungen dominieren auch tatsächlich in Ribeiros Interpretationsrichtung: *Kohlepapier* dominiert im Rohstoffsinne deutlich vor *Kohlenpapier* und zählbares *Kohleneimer* dominiert deutlich vor *Kohleeimer*. Aber wie immer gibt es Gegenbelege: So dominiert zwar *Kohlenkeller*, aber das bedeutungsähnliche *Kohlenschuppen* ist nur minimal häufiger als *Kohleschuppen*. Auch dominiert *Kohleofen* vor *Kohlenofen*, aber *Holzkohlegrill* ist etwa gleichauf mit *Holzkohlengrill*. Das heißt: Aus der *Kohlenkeller*-Dominanz können wir keine *Kohlenschuppen*-Dominanz ableiten und aus der *Kohleofen*-Dominanz keine *Holzkohlegrill*-Dominanz.

Zudem werden besonders Lerner die „richtige“ Interpretation schwierig finden: Meint *Kohleverbrauch* den Verbrauch des Rohstoffs Kohle oder sollte man nicht *Kohlenverbrauch* sagen, weil es um den Verbrauch zählbarer Kohlen geht? Kann man die Zählbarkeit, die nach Ribeiro *Kohlen* erwarten lässt, zum Beispiel an einer Mengenangabe erkennen?

*Gemäß dieser Prognose steigt der **Kohlenverbrauch** von 3,1 Mrd. Tonnen im Jahr 1990 auf mehr als 5,8 Mrd. Tonnen im Jahr 2020.*

(Salzburger Nachrichten 1993, IDS-Korpora)

Offenbar kann man das nicht. Vielleicht legt die Mengenangabe sogar eher *Kohle* nahe?

*Allein der chinesische **Kohleverbrauch** wird von zuletzt rund 2,2 Milliarden Tonnen im Jahr wohl auf rund vier Milliarden Tonnen im Jahr 2020 anwachsen.*

(Die Zeit 2007, IDS-Korpora)

Und geht es im Beleg aus der Tiroler Tageszeitung nicht doch um den Rohstoff, lässt also nach Ribeiro *Kohle* erwarten?

*Die USA bekannten sich somit erstmals zu bindenden Zielen und sprachen sich für eine global geltende Obergrenze für Kohlendioxid-Ausstöße aus dem Öl- und **Kohlenverbrauch** aus.*

(Tiroler Tageszeitung 1996, IDS-Korpora)

Und wie schließlich sieht es mit dem semantischen Leitkriterium im letzten *Kohlen*-Beleg aus, der innerhalb eines kleinen Abschnitts zwischen *Kohle* und *Kohlen* variiert?

*Ich glaube, Bush weiß nicht, worum es geht. Die Regierung redet über China und dessen Probleme mit der **Kohlefeuerung**. Aber China hat seit 1996 seinen **Kohlenverbrauch** um 24 Prozent verringert. Wenn die USA das geschafft hätten, hätten wir wahrscheinlich jetzt schon unsere Kioto-Verpflichtungen erfüllt.*

(die tageszeitung 2001, IDS-Korpora)

Prüfen wir eine weitere „semantische Differenzierung“: Ribeiro (1996, S. 16) liest *Kirschen* in *Kirschenaugen* oder *Kirschenmund* als Signal für eine „Vergleichsrelation“. Dagegen sprechen wiederum die zahlreichen Belege aus den Korpora der geschriebenen Sprache des IDS. *Kirschen* ist keineswegs überall Vergleichsmarker und *Kirsch* steht auch in Zusammensetzungen mit Vergleichsrelation. Auch hier sind die Varianten von ihrer Bedeutung her nicht zuverlässig zu erklären.

*Wildschweinbraten mit **Kirschensauce***

(Die Zeit 2001, IDS-Korpora)

*Den Flan im auf 200 Grad vorgeheizten Ofen auf der untersten Rille während 45 bis 50 Minuten backen. Sollte die Oberfläche zu rasch bräunen, mit Alufolie abdecken. Am besten schmeckt der **Kirschenflan** lauwarm. Nach Belieben mit etwas Puderzucker bestäuben.*

(Zürcher Tagblatt 1997, IDS-Korpora)

*Mulans Gesicht etwa entspricht mit zierlichen Zügen und **Kirschmund** dem chinesischen Schönheitsideal.*

(die tageszeitung 1999, IDS-Korpora)

*Der Mann provoziert, schmeichelt, droht, lockt. Die Frau wehrt ab und kommt doch immer näher. Der rote **Kirschmund** öffnet sich, in weisser Unschuld bebt das Unterhemd.*

(Zürcher Tagesanzeiger, IDS-Korpora)

In ihrer Bedeutung klarer zu unterscheiden sind dagegen einige etablierte Varianten. Linguisten haben solche etablierten Varianten zusammengetragen – mühsam, denn diese Art sichtbarer Bedeutungsvarianz ist selten.

Hausfront versus **Häuserfront**

Volkskunde versus **Völkerkunde**

Staatsgemeinschaft versus **Staatengemeinschaft**

Landeskonzferenz versus **Länderkonferenz**

(Ramers 1997, S. 35)

Gasthaus versus **Gästehaus**

(Fleischer/Barz 1995, S. 140)

Mitunter können wir hier einen Sinn herauslesen. *Hausfront* bezeichnet die Front eines Hauses, *Häuserfront* bezeichnet eine Front aus mehreren Häusern.

*Ben's Chili Bowl, dieses kleine Soulfood-Lokal an der Grenze zum Nordosten, ist inzwischen ein Wahrzeichen. Seit 1958 steht es fast unverändert da, die **Hausfront** noch per Hand bemalt: links ein Hotdog, rechts ein dicker Burger.*

(Die Zeit 2007, IDS-Korpora)

*Links die Kirche St. Alban, die noch aus karolingischer Zeit stammt. Rechts die niedere **Häuserfront** Kleinbasels. Wieder links die Pfalz mit dem romanischen Münsterchor.*

(Die Zeit 2010, IDS-Korpora)

Aber bei *Gasthaus* versus *Gästehaus* kommen wir mit Bedeutungs differenzierung nicht weiter: Beide bezeichnen ja ein Haus, das Gäste beherbergt. Vielleicht ist *Gästehaus* ohnehin nur in Abgrenzung zu früherem *Gasthaus* gebildet worden, „einem wort vom höchsten alter zugleich und noch jetzt in voller geltung“ (Deutsches Wörterbuch 1878, S. 1497). Der Plural in *Gästehaus* macht

also lediglich darauf aufmerksam, dass wir es nicht mit *Gasthaus* im schon lange etablierten Sinne zu tun haben. Vielzahl soll hier nicht betont werden.

Plurale sind also oft markierte Formen. Markiertheit sehe ich als besonderes Ereignis; entsprechend klein erwarte ich den prozentualen Anteil pluraler Erstglieder im Vergleich zu ihren singularen Normalvarianten. Einige von mir recherchierte Erstgliedvarianten aus den DeReKo-Schnittstellenanalysen zeigen dieses erwartete Verhältnis. In Tabelle 30 steht die Pluralform im Hintergrund:

<i>Dorf</i> <i>Dörfer</i>	189 744 195
<i>Front</i> <i>Fronten</i>	27 330 661
<i>Options</i> <i>Optionen</i>	4 441 1 237
<i>Wort</i> <i>Wörter</i>	77 798 8 790

Tabelle 30

In Tabelle 31 allerdings nicht:

	DeReKo	Google
<i>Aortaklappe</i> <i>Aortenklappe</i>	13 85	5 140 34 300
<i>Themawechsel</i> <i>Themenwechsel</i>	95 471	24 900 37 000

Tabelle 31

Liegt das am Singularauslaut auf *a*, den *Aorta* und *Thema* gemeinsam haben? Offenbar nicht, denn andere auf *a* auslautende Erstglieder neigen wiederum zum Singular.

	DeReKo	Google
<i>Ligaausschuss</i> <i>Ligenausschuss</i>	1 776 32	72 700 127
<i>Mensabetrieb</i> <i>Mensenbetrieb</i>	66 17	4 840 113

Tabelle 32

Nur vorsichtig formuliere ich deshalb: Mitunter geben wir Erstgliedern durch ihre Pluralform eine herausgehobene Vielzahlbedeutung. Spuren dafür, dass uns Vielzahl anzeigende Pluralvarianten als etwas Herausgehobenes bewusst sind, finden sich auch in unserer IDS-Korrespondenz.

Mail vom 21.6.2006

Sehr geehrte Frau Trabold,

von einem Kunden aus der Maschinenbaubranche habe ich die Anfrage, ob es „Spanmenge“ oder „Spänmenge“ heißt. Er möchte seine Prospektmaterialien einheitlich gestalten und nur noch eines der Worte verwenden. Grundsätzlich denke ich, es gibt die Möglichkeit, beides zu verwenden. Als Linguist tendiere ich allerdings zu Spanmenge, da Späne ja schon eine Menge ist. Liege ich richtig?

Über Antwort freue ich mich und verbleibe jetzt schon mit einem herzlichen Dankeschön

A. F.-G.

Dazu können wir uns natürlich mehr denken als der linguistische Anfrager andeutet: Stimmt denn, dass *Späne* schon eine Spanmenge bezeichnet und wir deshalb keine *Spänmenge* brauchen? Dann haben wir eine Menge von etwas, das Span ist. Oder verlangt *Menge* geradezu nach Vielzahl? Sollte es also *Fischemenge*, *Flüssigkeitenmenge*, *Geldermenge*, *Hotelsmenge*, *Ölemenge*, *Schnäpsemenge*, *Salzmenge*, *Stoffmenge* und eben auch *Spänmenge* heißen? Eine Menge kann ja nicht aus einem Span bestehen.⁸⁶ Und wenn beides geht: Bedeuten *Spanmenge* und *Spänmenge* dasselbe?

Im Juli 2010 habe ich rund 1 220-mal *Spanmenge* und 490-mal *Spänmenge* gegoogelt.⁸⁷ Einen die Vielzahl betreffenden Bedeutungsunterschied konnte ich bei Durchsicht der Belege nicht erkennen.

⁸⁶ So wie Semmelknödel nicht aus einem Semmel. Jedenfalls wenn es nach Karl Valentin geht:

„Man sagt von jeher Semmelknödel.

Ja, zu EINEM – aber zu mehreren Semmelknödel sagt man **Semmelknödeln**. Semmel ist die Einzahl, das mußt Ihnen merken, und Semmeln ist die Mehrzahl.“

(Valentin 2006, S. 83)

⁸⁷ Bei einem pluralen Zweitglied nähert sich das Verhältnis übrigens an: *Spanmengen* hat Google 1 370-mal geliefert, *Spänmengen* immerhin 902-mal. Kurz nachgegangen bin ich meinem Eindruck, das Erstglied sei vielleicht je nach Zweitglied pluralisch markiert. Diese Idee kam mir bei folgender Einzelrecherche: *Hauswand* ist 3 029-mal belegt, *Häuserwand* 218-mal; *Hauswände* ist 817-mal belegt, *Häuserwände* 463 mal. Beim singularen Zweitglied *Wand* sind also 93,3% singulare und 6,7% plurale Erstglieder; beim pluralen Zweit-

Die Geschwindigkeit des Spanstromes wird so gesteuert, dass die im Pufferraum befindliche **Spanmenge** möglichst konstant bleibt.

(www.wikipatents.com, 30.1.1975)

Die Austragung der Späne aus dem Silo muss dem Gewicht der zu lagern-
den **Spänmenge** angepasst sein.

(www.umweltschutz-bw.de, Juli 2010)

<i>Abgabemenge</i>	5
Abgabemenge	122
Blättermenge	3
<i>Blattmenge</i>	2
Büchermenge	6
<i>Buchmenge</i>	1
<i>Früchtemenge</i>	5
Fruchtmenge	13
Kohlenmenge	13
<i>Kohlemenge</i>	5
<i>Lichtermenge</i>	1
Lichtmenge	167

Tabelle 33

Auch andere Belege mit *Menge* sprechen für Gleichheit. Ich habe wieder exemplarisch einige Varianten in den IDS-Korpora der geschriebenen Sprache gesucht. In Tabelle 33 dominiert meist eine der beiden Erstgliedformen, mitunter deutlich: Bei *Abgabe-* und *Lichtmenge* dominieren die singularen Erstglieder, bei *Büchermenge* die pluralen.

Allerdings entscheidet Google im September 2010 zugunsten *Blattmenge* mit 6090 Belegen gegenüber 994 Belegen für *Blättermenge*. Und auch im November 2010 stehen 6450 Belege für *Blattmenge* nur 1110 Belegen für *Blättermenge* gegenüber. Soll ich Google mehr vertrauen als den DeReKo-Schnittstellenanalysen beziehungsweise den Korpora der geschriebenen Sprache des

glied *Wände* dagegen erhöht sich der Anteil der pluralen Erstglieder auf 36,1%. Heißt das, dass bei pluralischem Zweitglied der Anteil der pluralen Erstglieder immer steigt? Mit dem singularen Zweitglied *Dach* wird *Häuser* nur zu 1,4% zusammengesetzt; beim pluralen Zweitglied *Dächer* erhöht sich der Anteil auf 13,4%. Das singulare Zweitglied *Laden* wird allerdings mit dem pluralen Erstglied *Bücher* zu 6,3% zusammengesetzt, beim pluralen Zweitglied *Läden* schrumpft der Anteil auf 2,5%.

IDS? Ist Google zuverlässiger, weil einfach größer? Dieselbe Frage stellt sich mir noch einmal bei Tabelle 50. Dort interessiert mich auch die Antwort. Hier geht es dagegen nicht um Zahlen, sondern um etwas davon Unabhängiges, nämlich die bedeutungsbedingte Varianz.

Schauen wir uns die Bedeutungen der Varianten in ihren natürlichen Kontexten näher an. Sind zum Beispiel *Blätter-* und *Blattmenge* oder *Bücher-* und *Buchmenge* gleich zu lesen?

*Wird, wie bei Rosenöl, traditionell destilliert, ergeben fünf Tonnen Rosenblätter gerade mal einen Liter Rosenöl. Bei der Extraktion mit Lösemitteln hingegen läßt sich aus der **Blättermenge** die zehnfache Menge gewinnen.*
(die tageszeitung 1994, IDS-Korpora)

*Ab einer **Blattmenge** von dreißig Blättern nahm die Biomasse der darin verborgenen Wirbellosen sogar exponentiell zu.*
(spektrum direkt 2004, IDS-Korpora)

*Er beschleunigte die bekannten Handgriffe, und in den nächsten drei Stunden hatte er fast das Doppelte der üblichen **Büchermenge** verpackt und die Hälfte des mittleren Tisches völlig aufgearbeitet.*
(Paarmann 2000, IDS-Korpora)

*In der Pariser Nationalbibliothek geht die Individualität des einzelnen Buches im Prozeß der Katalogisierung in der Allgemeinheit der **Buchmenge** unter.*
(die tageszeitung 1998, IDS-Korpora)

Das sieht sehr gleich aus. Andererseits erhellen sich uns etwa bei *Lichtmenge* neben *Lichtermenge* durchaus Unterschiede; einmal ist es die Menge von Licht als Substanz, einmal ist es die Menge vieler einzeln wahrnehmbarer, festlicher Lichter.

*Ein Grossteil der Birnen sei in den letzten zwei Jahren bereits ausgewechselt worden, sagt René Dürr von der Elektra Gaiserwald. Bis anhin seien Quecksilberdampf-Birnen verwendet worden, jetzt würden diese durch Natriumdampf-Birnen ersetzt. Diese brauchen weniger Strom und leuchten die Strasse erst noch besser aus. «Die neuen Birnen spenden fast die doppelte **Lichtmenge**, und das orange-gelbe Licht hebt vor allem bei Regen die Konturen und Farben besser hervor», sagt Dürr.*
(St. Galler Tagblatt 1997, IDS-Korpora)

Denn der 65-jährige Pensionär ist schon seit 42 Jahren Kärlichs St. Martin, damit wohl der Dienstälteste der Verbandsgemeinde Weißenthurm.

*Und die Pferde, die ihn im festlichen Umzug durch den Ort getragen haben, sind zahlreich. „Zehn bis zwölf Pferde werden es wohl sein. Alle Namen kriege ich schon gar nicht mehr zusammen“, sagt Scheid. Aber an Bento, das erste Pferd, auf dem er als barmherziger Mann in der **Lichtermenge**⁸⁸ ritt, kann er sich noch gut erinnern.*

(Rhein-Zeitung 2003, IDS-Korpora)

So weit die Betonung von Vielzahl. Wie sieht es in den DeReKo-Schnittstellenanalysen mit der Betonung von Zugehörigkeit aus?

Eine spezielle Art von Zugehörigkeit ist Verwandtschaft. Eine spezielle Art Variante betont diese Zugehörigkeit: Bezeichnungen für Personen wie *Bürger*, *Lehrer* und *Pfarrer*, die in allen anderen Zusammensetzungen regelmäßig in ihrer Nullform erscheinen (*Bürgerbegehren*, *Lehrerzimmer*, *Pfarrerausbildung*),⁸⁹ verwenden wir seit Alters her in ihrer Genitiv-s-Form, sobald wir sie mit Bezeichnungen für verwandte Personen wie *Sohn* oder *Tochter* zusammensetzen.⁹⁰ Dieses spezielle Phänomen, das in der linguistischen Literatur, besonders der didaktischen, häufig herausgestellt wird, zeigt sich auch bei den DeReKo-Schnittstellenanalysen der Tabelle 34.

Bürger	1 488 038	Bürgers	570
<i>Bürgerhaus</i>	101 506	<i>Bürgershaus</i>	2
<i>Bürgertochter</i>	73	<i>Bürgerstochter</i>	247
Lehrer	49 830	Lehrers	151
<i>Lehrerzimmer</i>	2 427	<i>Lehrerszimmer</i>	0
<i>Lehrertochter</i>	17	<i>Lehrerstochter</i>	48
Nachbar	73 416	Nachbars	1 759
<i>Nachbarhaus</i>	3 648	<i>Nachbarshaus</i>	10
<i>Nachbarjunge</i>	198	<i>Nachbarsjunge</i>	379
Pfarrer	2 016	Pfarrers	1 690
<i>Pfarrerwohnung</i> ⁹¹	25	<i>Pfarrerswohnung</i>	15
<i>Pfarrerfrau</i>	0	<i>Pfarrersfrau</i>	262

Tabelle 34

⁸⁸ *Lichtermenge* ist in den IDS-Korpora der geschriebenen Sprache nur vereinzelt belegt; auch Google hat im August 2010 nur 62 Belege dazu. Alle Belege bewegen sich aber ebenso im Kontext von Weihnachtsbäumen, Himmelszelten und demonstrierenden Menschenmengen.

⁸⁹ Von der Bedeutung her kann man hier auch Plural annehmen: *Begehren der Bürger*, *Zimmer für Lehrer*.

⁹⁰ Hier geht es, unserer Kultur gemäß, natürlich um Verwandtschaft im weiteren Sinne: Mit 65 Belegen ist auch *Nachbarshund* vertreten.

In Tabelle 34 ist gut zu sehen, dass die von mir recherchierten Personenbezeichnungen überwiegend die Nullform haben, wenn sie mit Bezeichnungen von Nichtpersonen zusammengesetzt sind, und dass sie die Genitivform haben, wenn sie mit anderen Personenbezeichnungen zusammengesetzt sind. So hat zum Beispiel *Bürgerhaus* an allen Zusammensetzungen mit *Bürger* einen relativ hohen Anteil von 6,8%, *Bürgershaus* dagegen an allen Zusammensetzungen mit *Bürgers* einen extrem geringen von 0,3%. Und andersrum hat *Bürgertochter* an allen Zusammensetzungen mit *Bürger* einen extrem geringen Anteil von 0,004%, während *Bürgerstochter* 43,3% aller Zusammensetzungen mit *Bürgers* ausmacht.

Bei *Lehrers* gibt es sogar nur einen einzigen Beleg für eine Zusammensetzung mit einem nichtpersonalen Zweitglied: *Lehrershand*.

*„Ich darf nicht in der Reie reden, vor allem wen ich in der Pause Genug Zeit hatte, Dume geschichten zu erzellen.“ – „Ich werd in der Pause keinen Unsinn erzählen oder Unnordnung schaffen, so das die Aufsicht einschreiten muss.“ – „Ich darf im Geschichtsunterricht nicht mit Papnasen herumalbern und meine Kammradn zersträuen.“ Und immer so weiter und jeden von diesen Sätzen hundert Mal im Präsens, Präteritum und Futur variiert; und dann jeweils dick rot unterkringelt und von garstiger **Lehrershand** kommentiert: „Fürchterliche Rechtschreibung! Deine Eltern können stolz auf dich sein! Du stehst schon mit einem Bein im Gefängnis!“*

(Berliner Zeitung 2004, IDS-Korpora)

Wie in Tabelle 35 aber zu sehen ist, kann daraus keine Regel abgeleitet werden, denn wir tendieren auch hier mal zu diesem, mal zu jenem.

Bürgersohn	198	<i>Bürgerssohn</i>	35
<i>Bürgertochter</i>	73	Bürgerstochter	247
Lehrersohn	145	<i>Lehrerssohn</i> ⁹²	35
<i>Lehrertochter</i>	17	Lehrerstochter	48
<i>Nachbarsohn</i>	12	Nachbarssohn	108
<i>Nachbartochter</i>	11	Nachbarstochter	191
<i>Pfarrersohn</i>	100	Pfarrerssohn	443
<i>Pfarrertochter</i>	0	Pfarrerstochter	474

Tabelle 35

⁹¹ Das Erstglied *Pfarr* wie in *Pfarrhaus* geht auf *Pfarre* (griechisch *parochia*) zurück. *Pfarr* ist also keine Variante zu *Pfarrer*.

So tendieren wir zwar wie beschrieben bei Zusammensetzungen von Personenbezeichnungen zur Genitivform, auffälligerweise ist aber bei Zusammensetzungen mit dem Zweitglied *Sohn* eine Tendenz zur Nullform zu erkennen, jedenfalls bei manchen Erstgliedern, nämlich bei *Bürger* und *Lehrer*. Dagegen entsprechen Zusammensetzungen mit *Sohn* als Zweitglied bei den Erstgliedern *Nachbar* und *Pfarrer* wieder der beschriebenen Tendenz zur Genitivform. Wie sollen sich Lerner sowas merken können? Hier würde nur ein Wörterbuch helfen, in dem man einzelne Erstglieder nachschlagen kann. Online ist sowas durchaus denkbar.

Mitunter verwenden wir also flexivisch markierte Erstglieder, um Vielzahl oder Zugehörigkeit auszudrücken. Wir können das tun, müssen es aber offenbar nicht und tun es auch nicht regelmäßig. Bedeutung lässt sich eben nicht immer in Regeln festschreiben: So wird im Kontext von *Blattmenge* eine pluralische Zahl von dreißig Blättern genannt. Muss deshalb das Erstglied nicht noch einmal pluralisch markiert werden? Oder sollte es gerade deshalb pluralisch markiert werden?

Und noch eine letzte bedeutungsbedingte Variante: Das Bedeutungsprinzip greift auch, wenn „Homonymendifferenzierung indiziert“ werden soll (Fleischer/Barz 1995, S. 137). Solche Fälle sind selten und werden von Linguisten mühsam zusammengetragen.

Heidekraut versus **Heidenangst**

(Fleischer/Barz 1995, S. 137)

Geistesgegenwart versus **Geisterstunde**

Landsmann versus **Landmann**

(Eichinger 2000, S. 51)

Die DeReKo-Schnittstellenanalysen zeigen 339-mal *Geisterwelt* und 338-mal *Geisteswelt* sowie 1439-mal *Landmann* und 28935-mal *Landsmann*. Womit die Auftretenshäufigkeit zum Beispiel von *die Heide* und *der Heide* belegt wird; eine Vielzahl-Zugehörigkeits-Variante liegt nicht vor. Wir signalisieren mit der Variante nämlich einen Bedeutungsunterschied, der schon den Erstgliedern eigen ist.

⁹² Gegen die Theorie, dass zum Beispiel *Lehrersohn* häufiger ist als *Lehrerssohn*, weil gedoppeltes *s* vermieden werden soll, spricht auch die Dominanz von *Nachbarssohn* und *Pfarrerssohn*.

Euphoniebedingte Varianten

Die Euphonieregel führt zur „rhythmischen Optimierung“ von Zusammensetzungen (Nübling/Szczepaniak 2009, S. 203). Nach ihr gestalten wir Erstglieder so, dass sie für unsere deutschen Ohren schön klingen, zum Beispiel trochäisch. Das trifft jedenfalls auf Varianten zu wie *Beeren* neben *Beer*. *Beere* als Erstglied wird in der trochäischen Pluralform *Beeren* verwendet, zum Beispiel in *Beerenschnaps*. Ist *Beere* aber in ein komplexes Erstglied eingebunden wie in *Blaubeere*, wird meist die trochäische Tilgungsform *Blaubeer* verwendet, zum Beispiel in *Blaubeerschnaps*.

Beeren + Zweitglied

Blaubeer + Zweitglied

Ausnahmen sind selten. *Blaubeerenschnaps* findet sich in den IDS-Korpora der geschriebenen Sprache gar nicht⁹³ und auch andere Zusammensetzungen mit *Blaubeeren* nur vereinzelt, zum Beispiel 4-mal *Blaubeerenwald* und ein mal *Blaubeerensuppe*.⁹⁴

In vielen von mir recherchierten Fällen verhalten sich Erstglieder so. Häufig variiert ein trochäisch zugeschnittenes Erstglied wie *Beeren*, *Hunde*, *Lieder* und *Meeres*, sobald es in ein komplexes Erstglied eingebunden ist. Die Erstgliedvariante wird trochäisch zugeschnitten, zum Beispiel *Erdbeer*, *Seehund*, *Volkslied* und *Schwarzmeer*, eher nicht *Erdbeeren*, *Seehunde*, *Schwarzmeeres* und *Volkslieder*.

⁹³ Was bekanntlich nicht wirklich aussagekräftig ist, denn „Korpusbelege können nur positiv interpretiert werden. Wenn ein Phänomen in einem Korpus belegt ist, dann existiert es auch. Wenn ein Phänomen in einem Korpus nicht belegt ist, kann man keine Aussage folgern“ (Perkuhn/Belica 2006, S. 4). Wie zudem nicht anders zu erwarten, hat Google im September 2010 immerhin 54-mal *Blaubeerenschnaps*, fragt mich aber fürsorglich: „Meinten Sie: *Blaubeerschnaps*?“, was zur fünffachen Belegmenge führt.

⁹⁴ *Der Zufall oder ein unseren Wünschen gehorsamer Zufall brachte es mit sich, daß Matzerath am Abend des soeben beschriebenen Badetages – wir aßen **Blaubeerensuppe** und hinterher Kartoffelpuffer* (Grass, Blechtrommel 1962, IDS-Korpora). Auch *Erdbeerenparfait*, *Erdbeerensulze* und *Erdbeerenragout* finden sich dort nur vereinzelt: *Am Freitag, 25. September; ist es soweit, dann wird von der Entenbrust über Tartarenbrötli bis zum **Erdbeerparfait** alles zu den 97er Weinen der Genossenschaft serviert* (St. Galler Tagblatt 1998, IDS-Korpora).

<i>Beeren-schnaps</i>	6	<i>Beer-schnaps</i>	0	<i>Erdbeer-schnaps</i> <i>Erdbeeren-schnaps</i>	6 0
<i>Hunde-bestand</i>	41	<i>Hund-bestand</i>	0	<i>Seehund-bestand</i> <i>Seehunde-bestand</i>	39 1
<i>Lieder-sammler</i>	18	<i>Lied-sammler</i>	1 ⁹⁵	<i>Volkslied-sammler</i> <i>Volkslieder-sammler</i>	33 8
<i>Meeres-flotte</i>	5	<i>Meer-flotte</i>	0	<i>Schwarzmeer-flotte</i> <i>Mittelmeeres-rand</i>	8 953 0

Tabelle 36

Verhalten sich Erstglieder, die von ihrer Gestalt her *Beeren*, *Hunde*, *Lieder* und *Meeres* ähnlich sind, genauso? Handelt es sich wenigstens hier um eine Regel oder doch zumindest um eine starke Tendenz? Ich tabelliere vier analoge Fälle aus den IDS-Korpora der geschriebenen Sprache.

Alle simplizischen trochäischen Erstglieder der Tabelle 37 dominieren wiederum mehr oder weniger deutlich; simplizische Nichttrochäer sind dabei gar nicht oder jedenfalls viel seltener belegt. Bei den komplexen Erstgliedern dominieren allerdings nicht durchgängig die trochäisch zugeschnittenen Formen: Bei *Birne* zum Beispiel und bei *Pferd* überwiegen nicht die Varianten *Birn* und *Pferd*, sondern wie bei den simplizischen Erstgliedern *Birnen* und *Pferde*.

In der letzten Zelle der Tabelle 37 dominiert *Jahres* vor *Jahr*. Hier konkurrieren auch ohnehin vor allem andere Erstgliedformen, nämlich *Jahres* mit *Jahrs*. Dabei ist *Jahres* die Form simplizischer Erstglieder, *Jahrs* die Form komple-

⁹⁵ Konrad Mautner, *Wiener Großindustrieller, Volkskundler, Liedsammler und Trachtenexperte*, starb am 15. Mai 1924 in Wien (Salzburger Nachrichten 1998, IDS-Korpora).

Birnen- <i>saft</i>	179	Birn- <i>saft</i> ⁹⁶	1	Hirschbirn- <i>saft</i>	1
				Mostbirnen- <i>saft</i>	3
				Graubirnen- <i>saft</i>	1
Pferde- <i>zucht</i>	230	Pferd- <i>zucht</i> ⁹⁷	3	Flusspferd- <i>zucht</i>	3
				Reitpferde- <i>zucht</i>	6
Bilder- <i>sammlung</i>	465	Bild- <i>sammlung</i>	9	Lichtbild- <i>sammlung</i>	4
				Lichtbilder- <i>sammlung</i>	1
Jahres- <i>budget</i>	2 319	Jahr- <i>budget</i>	0	Wahljahr- <i>budget</i>	1
				Vorjahres- <i>budget</i>	68

Tabelle 37

xer Erstglieder. Wiederum werden Trochäen gebildet: **Jahrestagung** neben *Frühjahrestagung*. Nur ausnahmsweise ist auch das simplizische Erstglied *Jahrs*, zum Beispiel in *Jahrsbeginn*.⁹⁸ Auch bei *Tages* neben *Tags* ist *Tags* als

⁹⁶ Nicht zu vergessen das passende Gedicht dazu: „Bodder, das is so Geschlodder; Kees, das is so Gefräß, aber **Birnsaft** – das gibt Kraft“, schmunkelt Erich Doss zu der „Bimbeshymne“ (Rhein-Zeitung 2000, IDS-Korpora). In Google habe ich im September 2010 immerhin 1 290 Belege für *Birnsaft* gefunden; *Birnensaft* ist dort mit über 30 000 Belegen zu finden.

⁹⁷ Alle drei Belege stammen aus der Rhein-Zeitung, zum Beispiel: *Drei Deckhengste stehen für die **Pferdezucht** im Kreis zur Verfügung* (Rhein-Zeitung 1996, IDS-Korpora).

⁹⁸ Denn zum Ärger der New Yorker UNO-Kasse überweist Bonn seinen zum **Jahrsbeginn** feststehenden Gesamtbeitrag immer nur in zwei Halbjahrestranchen und trug damit auch schon in den letzten Jahren zu den zeitweisen Engpässen bei (die tageszeitung 1993, IDS-Korpora); Am vergangenen Freitag sei der Euro unter das bisherige **Jahrstief** 2005 gefallen, am Dienstag unter das Jahrestief 2004 bei 1,1760 Dollar (Frankfurter Allgemeine Zeitung 2005, IDS-Korpora). Bei den rund 650 Belegen zu *Jahrs* mag es sich teilweise um Verschreibungen handeln.

simplizisches Erstglied vergleichsweise selten;⁹⁹ die Schnittstellenanalysen liefern überwiegend komplexe Erstglieder wie in *Alltagssprache* und *Bezirkstagspräsident*.

<i>Tages-</i> <i>form</i>	3 415	<i>Halbtags-</i> <i>form</i>	14
		<i>Halbtages-</i> <i>form</i>	1
<i>Tages-</i> <i>hektik</i>	22	<i>Alltags-</i> <i>hektik</i>	446
		<i>Alltages-</i> <i>hektik</i>	0

Tabelle 38

Auch hier gibt es also keine klaren Präferenzen, aus denen sich Regeln und Empfehlungen ableiten ließen.

Strukturierungsbedingte Varianten

Nach der Strukturierungsregel helfen wir unseren Hörerlesern bei der Strukturierung komplexer Zusammensetzungen. Linguisten haben dafür einige Beispiele zusammengetragen. Es sind mal Zusammensetzungen mit gleichem, mal Zusammensetzungen mit unterschiedlichem Zweitglied.

Stückzahl neben *Frühstücksei*

(Michel 2009, S. 337)

Werkzeug neben *Handwerkszeug*

Turmuhr neben *Kirchturmsuhr*

Hofmauer neben *Friedhofsmauer*

(Fuhrhop 1998, S. 191)

Marktbude neben *Jahrmarktsbude*

Steinbrocken neben *Gesteinsbrocken*

(Kürschner 2003, S. 53)

⁹⁹ *Der Obstbau, die Bienenzucht und „werkeln“ in seiner Werkstatt füllen den Tagsablauf des über 90jährigen Heinrich aus* (Vorarlberger Nachrichten 1998, IDS-Korpora). Auch hier ist mit Verschreibungen zu rechnen, vielleicht auch mit bewussten Spielereien: *Nach Dienstschluss kriecht sie dann in das kleine Zimmer, in dem Adrián und seine Cousine das Bett abwechselnd nutzen: er, der Nachtarbeiter, tagsüber, sie, die Tagsarbeiterin, nachts* (die tageszeitung 2004, IDS-Korpora).

Zeitreise neben *Hochzeitsreise*
Sichtblende neben *Ansichtskarte*
Fahrtrichtung neben *Abfahrtszeit*
 (Wegener 2003, S. 445)

Wo wir die Sollbruchstelle einer Zusammensetzung sehen, wählen wir eine Wortform (*des Frühstücks*, *des Gesteins*) oder fügen ein unparadigmatisches Fugen-s an: *Zeit* + s, *Fahrt* + s. Die einzige Erstgliedform, die in all diesen Beispielen verwendet wird, ist die paradigmatische oder unparadigmatische s-Form.¹⁰⁰

Auch die DeReKo-Schnittstellenanalysen liefern Belege dafür, etwa für *Nachtzug* neben *Sommernachtstraum* und *Weihnachtszeit*. In den IDS-Korpora der geschriebenen Sprache, die den DeReKo-Schnittstellenanalysen zugrunde liegen, aber mehr Daten enthalten als diese, finden sich 276 Belege für *Schiffahrtsmuseum* neben nur 10 Belegen für *Schiffahrtmuseum*. Auch andere komplexe Erstglieder mit *Fahrt* zeigen eine klare Tendenz zur s-Form: Nur 9 Belege für *Jugendwohlfahrtgesetz* stehen 102 Belegen für *Jugendwohlfahrtsgesetz* gegenüber. Ob hier überhaupt oder sogar immer eine Strukturierungsabsicht vorliegt, können die Analysen natürlich nicht zeigen. Auch war eine rechnergestützt systematische Auszählung von simplizischen und entsprechenden komplexen Erstgliedern noch nicht möglich.

Gewohnheitsbedingte Varianten

Nach der Gewohnheitsregel gestalten wir Schnittstellen so, wie wir es oft genug gehört und gelesen, gesagt und geschrieben haben. Wir richten uns gerne nach etablierten Zusammensetzungen. Etablierung ist ein historischer Prozess; also sollten wir mit historisch gewachsenen, historisch bedingten Varianten rechnen. Etablierung ist ein Prozess zwischen uns allen; also ist mit Varianten zu rechnen, die sich daraus ergeben, dass wir keine homogene Masse sind. Wir variieren regional, individuell und gruppensprachlich. Zum Schluss geht es also um

- historisch bedingte Varianten
- regional bedingte Varianten
- individuell und gruppensprachlich bedingte Varianten

¹⁰⁰ Pluralische Wortformen werden nicht verwendet. Weil sie so vehement vielzählige Bedeutung signalisieren?

Gewohnheitsbedingte historische Varianten

Erstgliedvarianten sind uns schon aus althochdeutscher Zeit überliefert und offenbar haben wir keine Notwendigkeit gesehen, sie auf eine gemeinsame Form zu reduzieren. Wir leben also schon lange und gut mit Varianten. So variieren im Althochdeutschen nach Plank (1981, S. 38) zum Beispiel *erden* neben *ërde*, *ëren* neben *ēre* und *ēr* sowie *kirch* neben *kirchen*. Historisch „erklären sich [...] neuhochdeutsche Paare wie *Tagedieb*, *-lohn* – *Tageslicht* (althochdeutsch noch *tagalioht*)“ (Fleischer/Barz 1995, S. 136). Ortner et al. (1991, S. 94) weisen auf schon historische Formen mit *e*-Tilgung hin wie in *Birnbaum* neben *Birnensaft*, *Erbsbrei* neben *Erbsensuppe* und *Kirchturm* neben *Kirchenbesuch*.

Als Beispiel habe ich in den DeReKo-Schnittstellenanalysen vollständig alle echten Zweifelsfälle mit den Erstgliedvarianten *Kirchen* neben *Kirch* recherchiert. Die Recherche zeigt, dass „nicht selten Komposita mit jeweils mehreren Alternativen eines Determinans-Gliedes lexikalisiert werden“ (Plank 1981, S. 38). Plank spricht von „lexikalischer Tradierung“ (ebd.). Tabelle 39 lässt erkennen, dass beide Varianten „tradiert“ sind, denn Zusammensetzungen älteren Ursprungs haben mal die eine, mal die andere Variante, etwa *Kirchensprengel* und *Kirchacker*.

Dominanz von <i>Kirchen</i>		Dominanz von <i>Kirch</i>	
<i>Kirchenbesuch</i>	500	<i>Kirchacker</i>	65
<i>Kirchbesuch</i>	35	<i>Kirchenacker</i>	25
<i>Kirchenchor</i>	54 672	<i>Kirchgang</i>	2 408
<i>Kirchchor</i>	44	<i>Kirchengang</i>	56
<i>Kirchenfest</i>	910	<i>Kirchgänger</i>	1 125
<i>Kirchfest</i>	32	<i>Kirchengänger</i>	53
<i>Kirchengebäude</i>	756	<i>Kirchgeld</i>	346
<i>Kirchgebäude</i>	15	<i>Kirchengeld</i>	64
<i>Kirchenportal</i>	407	<i>Kirchhof</i>	5 802
<i>Kirchportal</i>	12	<i>Kirchenhof</i>	61
<i>Kirchensprengel</i>	50	<i>Kirchplatz</i>	11 017
<i>Kirchsprengel</i>	14	<i>Kirchenplatz</i>	3 822
<i>Kirchentreppe</i>	163	<i>Kirchturm</i>	10 409
<i>Kirchtreppe</i>	30	<i>Kirchenturm</i>	148
<i>Kirchentür</i>	564		
<i>Kirchtür</i>	18		

Tabelle 39

Einzelrecherchen in den IDS-Korpora der geschriebenen historischen Sprache bestätigen die Tendenzen der Tradition: Das Historische Korpus des IDS mit 45 Millionen laufenden Wörtern aus der Fachliteratur, der Belletristik und aus Zeitungen von 1700 bis 1918 tendiert in die allgemeine heutige Richtung. Ich habe einige der oben in den DeReKo-Schnittstellenanalysen recherchierten Tendenzen gegengetestet. Das Historische Korpus zeigt in Tabelle 40 ähnliche Zahlenverhältnisse wie die DeReKo-Schnittstellenanalysen.

Auch Einzelrecherchen in historischen Teilkorpora des IDS zeigen diese Tendenzen: So habe ich im Goethe-Korpus fünf Belege für *Kirchhof* und keinen für *Kirchenhof* gefunden. Dort gibt es einen Beleg für *Kirchengebäude* und einen für *Kirchenbesuch* ohne *Kirch*-Variante sowie einen für *Kirchgang* und 13 Belege für *Kirchhof* ohne *Kirchen*-Variante. Auch die Kinder- und Hausmärchen der Brüder Grimm haben *Kirchentreppe* ohne *Kirch*-Variante und *Kirchhof* ohne *Kirchen*-Varianten. Thomas Mann hat *Kirchenchor* und *Kirchturm* ohne Varianten. Die IDS-Korpora mit älteren Texten tendieren also ebenfalls zu den mehrheitlich gewählten Schnittstellen.

	Historisches Korpus
<i>Kirchenbesuch</i> <i>Kirchbesuch</i>	11 0
<i>Kirchengebäude</i> <i>Kirchgebäude</i>	6 0
<i>Kirchentür</i> <i>Kirchtür</i>	57 35
<i>Kirchenhof</i> <i>Kirchhof</i>	0 691
<i>Kirchenplatz</i> <i>Kirchplatz</i>	9 16
<i>Kirchenturm</i> <i>Kirchturm</i>	4 172

Tabelle 40

Historisch bedingt sind auch die für Verwirrung unter Nichtsprachhistorikern sorgenden, etablierten Zusammensetzungen mit heute nicht mehr gebräuchlichen Flexionsformen, etwa *Hahnenschrei* zu althochdeutsch *hanin* ‘des

Hahns'.¹⁰¹ So auch *Schwanenhals*. Solche Relikte sind allerdings nicht übermäßig häufig, also kein zu großes Problem für Lerner, Lehrer und Forscher. In den DeReKo-Schnittstellenanalysen finden sich lediglich 30 Belege für *Hahnen* und 210 Belege für *Schwanen*. Darunter sind zahlreiche Belege in historischen Eigennamen überliefert wie *Hahnen-Apotheke*, *Hahnengasse* und *Schwanenmühle*. Sie machen den Hauptteil der Belege aus: So ist *Hahnen-Apotheke* am Gesamt aller *Hahnen*-Belege mit 16 Belegen beteiligt (53,3%) und *Schwanen-Apotheke* am Gesamt aller *Schwanen*-Belege mit 100 Belegen (47,6%). Einfluss auf die Gestaltung haben sicher auch solche, vielleicht gerade solche Vorkommen.

Gewohnheitsbedingte regionale Varianten

Auch regional haben wir uns an etablierte Gestaltungen von Zusammensetzungen gewöhnt. Manchmal, ohne uns des Regionalen bewusst zu sein. So hat zum Beispiel, während im April 2010 die Aschewolke des isländischen Eyjafjalla die halbe Welt lahmlegte, ein aufmerksamer Sprachbeobachter bei uns nachgefragt:

Von: xy@telering.at
 An: sekretariatdeutscher-sprachrat.de
 Datum: Tue, 20 Apr 2010
 Betreff: Asche-Wolke??

Heißt es nicht richtig Aschenwolke?!
 So wie Nischenprodukt, Taschenträger ...
 MfG!

Der aufmerksame Sprachbeobachter geht also vor allem von seiner Gewohnheit aus: *Nischenprodukt* und *Taschenträger* sind ihm Vorbild. Und hat er nicht in seiner näheren Umgebung schon oft genug *Aschenwolke* gehört und gelesen?

Die DeReKo-Schnittstellenanalysen erlauben einen schnellen Überblick über regionale Varianz. Deshalb habe ich schnell recherchieren können, dass *Aschewolke* die allgemein üblichere Variante ist, aber Österreicher daneben zu 43,7% zu *Aschenwolke* tendieren.

¹⁰¹ Vgl. Paul (1975, S. 336) und Barz (2005, S. 723f.).

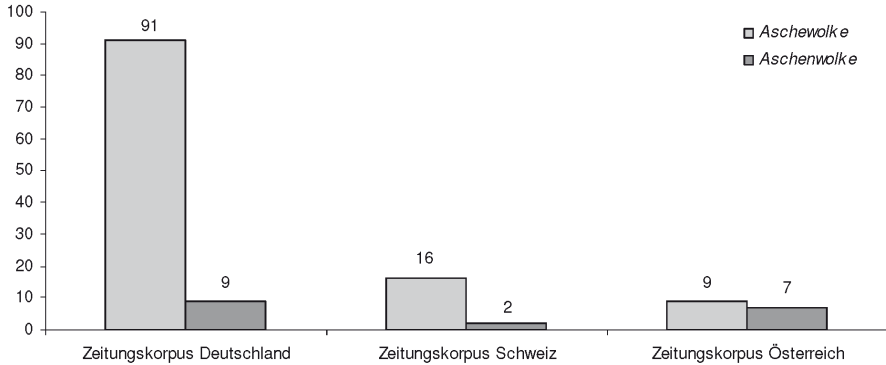


Diagramm 3

Linguisten, namentlich Dialektografen, weisen allenthalben auf regionale Varianten bei der Gestaltung von Zusammensetzungen hin. Vor allem werden österreichische Formen mit paradigmatischem und unparadigmatischem *s* gesammelt wie in *Unfallsgeschehen* und *Fabriksarbeiter*, die bundesdeutsche Nullformen wie in *Unfallgeschehen* und *Fabrikarbeiter* variieren.

Aufnahmsgesuch

(Ammon et al. 2004, S. 62)

Fabriksarbeiter, Unfallsgeschehen, Zugsunglück, Gepäcksstück

(Kellermeier-Rehbein 2005, S. 23)

Personsbeschreibung

(ebd., S. 99)

Auch das Deutsch der Schweiz hat nach Kellermeier-Rehbein (2005, S. 26) zusätzliche *s*-Formen wie in *Anleihensmarkt*, *Zugsunglück*, *Ausfuhrsartikel*, *Sportmeldung* und *Kirchsturm* neben bundesdeutschen Formen ohne *s* wie in *Sportmeldung* und *Kirchturm*.

In den DeReKo-Schnittstellenanalysen gibt es allerdings kein Indiz dafür, dass Sprecherschreiber deutscher Großregionen insgesamt zu bestimmten Gestaltungen von Zusammensetzungen neigen, etwa vermehrt zur *s*-Form. Wie die Statistik von Noah Bubenhofer in Tabelle 41 zeigt, werden die Schnittstellen insgesamt in allen drei deutschen Großregionen prozentual ähnlich gestaltet.

	Zeitungskorpus Deutschland ¹⁰²	Zeitungskorpus Schweiz ¹⁰³	Zeitungskorpus Österreich ¹⁰⁴
<i>0</i>	70,21%	72,09%	71,34%
<i>s</i>	16,71%	16,01%	15,92%
<i>es</i>	3,64%	2,51%	3,55%
<i>n</i>	6,25%	6,02%	6,19%
<i>en</i>	1,65%	1,92%	1,65%
<i>er</i>	0,93%	0,80%	0,75%
<i>e</i>	0,51%	0,56%	0,51%

Tabelle 41

Allenfalls ist zur Tabelle 41 zu bemerken, dass Österreicher und Schweizer gerade umgekehrt etwas weniger zu Erstgliedern mit *s* oder *es* neigen. Die Dialektografen haben dennoch auch Recht, denn der *s*- und Nullformengebrauch gleicht sich statistisch wieder aus: Gesammelt haben die Dialektografen nämlich auch österreichische Formen ohne *s* wie in *Adventkalender* neben bundesdeutschen Formen mit *s* wie in *Adventskalender* sowie österreichische Nullformen auf *e* wie in *Börseaufsicht* neben bundesdeutschen *en*-Formen wie in *Börsenaufsicht*. Auch im Deutschen der Schweiz finden sich *s*-lose Formen und Nullformen wie in *Börseaufsicht*. Hier einige für das Deutsch der Schweiz gesammelte Erstgliedvarianten.

Auslandhilfe, Bahnhofareal, Jahrzahl, Namenaktion, Hemdärmel
(Kellermeier-Rehbein 2005, S 26)

¹⁰² Das Zeitungskorpus Deutschland enthält: Berliner Morgenpost, Berliner Zeitung, Braunschweiger Zeitung, Computer Zeitung, Der Spiegel, Deutsche Presseagentur, Die Rheinpfalz, die tageszeitung, Die Welt, Die Zeit, Frankfurter Allgemeine, Frankfurter Rundschau, Hamburger Morgenpost, Hannoversche Allgemeine, Mannheimer Morgen, Neues Deutschland, Nürnberger Nachrichten, Nürnberger Zeitung, Rhein-Zeitung, spektrum direkt und VDI Nachrichten. Längerfristig soll das Zeitungskorpus Deutschland aufgedetailliert werden, denn die Unterschiede zum Beispiel zwischen der Berliner Zeitung und den Nürnberger Nachrichten interessieren ja durchaus. So zeigt das Zeitungskorpus Deutschland 15-mal das eigentlich für Österreich typische *Gansbraten*; alle Belege stammen aus den Nürnberger Nachrichten beziehungsweise der Nürnberger Zeitung.

¹⁰³ Das Zeitungskorpus Schweiz enthält: Die Südostschweiz, St. Galler Tagblatt und Zürcher Tagesanzeiger. Längerfristig ist vorgesehen, das Zeitungskorpus Schweiz dem bundesdeutschen Korpus umfänglich anzupassen.

Schafkäse, Gansbraten, Taglohn¹⁰⁵

(ebd., S. 97)

Adventkalender

(ebd., S. 99)

Börseaufsicht, Marmeladebrot

(ebd., S. 98)

Insgesamt mögen sich die Gestaltungen von Zusammensetzungen also gleichen, im Detail gibt es auffällige Unterschiede. Dazu habe ich in den DeRe-Ko-Schnittstellenanalysen einige in der Forschungsliteratur immer wieder genannte Varianten recherchiert.

	Zeitungs- korpus Deutschland	Zeitungs- korpus Schweiz	Zeitungs- korpus Österreich
<i>Adventkalender</i>	23	1	746
<i>Adventskalender</i>	4 127	475	23
<i>Anleihemarkt</i>	744	1	198
<i>Anleihenmarkt</i>	173	14	41
<i>Anleihensmarkt</i> ¹⁰⁶	0	0	0
<i>Aufnahmegesuch</i>	17	42	16
<i>Aufnahmsgesuch</i> ¹⁰⁷	0	0	1
<i>Auslandhilfe</i>	1	64	8
<i>Auslandshilfe</i>	412	21	235
<i>Bahnhofareal</i>	0	576	2
<i>Bahnhofsareal</i>	132	66	202

¹⁰⁴ Das Zeitungskorpus Österreich enthält: Burgenländische Volkszeitung, Die Presse, Kleine Zeitung, Neue Kronen-Zeitung, Oberösterreichische Nachrichten, Salzburger Nachrichten, Tiroler Tageszeitung und Voralberger Nachrichten. Längerfristig ist vorgesehen, das Zeitungskorpus Österreich dem bundesdeutschen Korpus umfänglich anzupassen.

¹⁰⁵ „Fleischer/Barz (1995, 141) hatten die fugenlose Verwendung von *Gans*- und *Tag*- als Bestimmungswörter noch ausschließlich dem österreichischen Deutsch zugeschrieben. Für *Gans*- trifft das immer noch zu, aber bei *Tag* ist dieses Bildungsmuster inzwischen auch in der Schweiz und teilweise in Süddeutschland gebräuchlich.“ (Kellermeier-Rehbein 2005, S. 97). Oder sind die Korpora jetzt einfach aussagekräftiger?

¹⁰⁶ Auch die Google-Suche ergab im August 2010 genau 0 Treffer.

¹⁰⁷ Auch *Aufnahmsgesuch* habe ich im September 2010 nur 65-mal googeln können.

<i>Börseaufsicht</i>	0	0	87
<i>Börsenaufsicht</i>	3 066	218	217
<i>Fabrikarbeiter</i>	563	76	44
<i>Fabriksarbeiter</i>	1	0	206
<i>Hemdärmel</i>	9	1	4
<i>Hemdsärmel</i>	93	5	16
<i>Kirchturm</i>	4 217	663	901
<i>Kirchsturm</i> ¹⁰⁸	0	0	0
<i>Jahrzahl</i>	2	116	21
<i>Jahreszahl</i>	2 189	221	539
<i>Marmeladebrot</i>	7	1	14
<i>Marmeladenbrot</i>	97	0	8
<i>Personsbeschreibung</i>	0	0	1 379
<i>Personenbeschreibung</i>	950	30	386
<i>Taglohn</i>	7	49	9
<i>Tagelohn</i>	38	4	4
<i>Zugunglück</i>	2 024	180	580
<i>Zugsunglück</i>	1	33	343

Tabelle 42

Gleich das zweite Beispiel der Tabelle 42 erstaunt: Kellermeier-Rehbein (2005, S. 26) haben ja *Anleihensmarkt* als typisch schweizerisch aufgeführt; es lässt sich aber kein Beleg dafür finden. Ebenso wenig wie für *Kirchsturm*. Und *Aufnahmsgesuch* ist auch nur einmal im österreichischen Korpus belegt. Kein oder nur ein Beleg muss nichts heißen, kann aber. Zumal auch Google in allen drei Fällen minimale Belegzahlen auswirft: *Aufnahmsgesuch* ist dort im November 2010 genau 74-mal belegt,¹⁰⁹ *Aufnahmegesuch* dagegen rund 16 300-mal.

¹⁰⁸ Auch die Google-Suche ergab nur einen Beleg in <http://www.weltreisebericht.com> am 13.1.2007, in dem allerdings nur beim ersten Vorkommen die *s*-Form erscheint, beim zweiten Vorkommen wurde die Form ohne *s* verwendet: *Höchster Kirchsturm der Welt? Das weiß auch jeder von Euch: Welches ist der höchste Kirchturm der Welt und wo steht er?*

¹⁰⁹ Wie das bei Google oft so ist, ist unter den Belegen auch ein Metabeleg aus Ammon et al. (2004). Google Books macht's möglich.

Andere von Dialektografen als typisch eingestufte Gestaltungen von Zusammensetzungen lassen sich dagegen geradezu ideal in den DeReKo-Schnittstellenanalysen nachweisen, etwa österreichisch *Adventkalender*, *Personsbeschreibung* und *Zugsunglück*. Das spiegelt sich auch in einzelnen Zeitungen wider. Recherchiert habe ich für Tabelle 43 zusätzlich zum gesamten Zeitungskorpus der Region jeweils eine große regionale Zeitung.

	D	FAZ	CH	Zür. Tagesanz.	A	Salzb. Nachr.
Adventkalender	23	0	1	0	746	61
Adventskalender	4 127	48	475	13	23	4
Fabrikarbeiter	563	48	76	13	44	13
Fabriksarbeiter	1	1	0	0	206	70
Personsbeschreibung	0	0	0	0	1 379	52
Personenbeschreibung	950	45	30	5	386	46
Zugunglück	2 024	101	180	54	580	133
Zugsunglück	1	0	33	2	343	64

Tabelle 43

Wie sieht es regional aus mit einzelnen Erstgliedern? Als Beispiel bin ich dem bei Kellermeier-Rehbein (2005, S. 97) als typisch österreichisch vorgestellten *Schafkäse* nachgegangen. Es ist tatsächlich im österreichischen Zeitungskorpus deutlich stärker vertreten als im bundesdeutschen und schweizerischen. Auch bei allen anderen echten Zweifelsfällen neigen Österreicher zur Nullform *Schaf*, Bundesdeutsche und Schweizer zur Genitivform *Schafs*.

	Zeitungskorpus Deutschland	Zeitungskorpus Schweiz	Zeitungskorpus Österreich
<i>Schafkäse</i>	27	15	254
<i>Schafskäse</i>	758	26	128
<i>Schafmilch</i>	21	28	52
<i>Schafsmilch</i>	72	17	19
<i>Schafwolle</i>	446	132	357
<i>Schafswolle</i>	234	7	13
<i>Schafffleisch</i>	126	32	94
<i>Schafsfleisch</i>	35	0	0

Tabelle 44

Allerdings neigen nach Tabelle 44 in einigen Fällen auch Bundesdeutsche zur Nullform, etwa in *Schafwolle* und *Schaffleisch*.¹¹⁰ Und insgesamt ist in allen drei Großregionen nach Tabelle 45 eine deutliche Neigung zur Nullform zu verzeichnen.

	Zeitungskorpus Deutschland	Zeitungskorpus Schweiz	Zeitungskorpus Österreich
Erstglied <i>Schaf</i>	6921	1 744	2 877
Erstglied <i>Schafs</i>	3 634	336	677

Tabelle 45

Das heißt: Es gibt regional bedingte Varianten. Manchmal gibt es feststellbare Tendenzen in der einen oder anderen Region, aber verlässlich gelten sie nur für einzelne Erstglieder, meist sogar nur für einzelne Erstglieder mit gleichem Zweitglied, wie das Erstglied *Schaf* neben *Schafs* mit dem Zweitglied *Fleisch*.

Gewohnheitsbedingt individuelle und gruppensprachliche Varianten

Gewohnheitsbedingt entstehen Varianten auch zwischen Individuen. Auch Individuen selbst variieren mitunter, auch innerhalb eines Textes. Mit einem Beleg für Varianten innerhalb eines Textes hatten wir es schon im Kapitel über bedeutungsbedingte Varianten zu tun.

*Ich glaube, Bush weiß nicht, worum es geht. Die Regierung redet über China und dessen Probleme mit der **Kohlefeuerung**. Aber China hat seit 1996 seinen **Kohlenverbrauch** um 24 Prozent verringert. Wenn die USA das geschafft hätten, hätten wir wahrscheinlich jetzt schon unsere Kioto-Verpflichtungen erfüllt.*

(die tageszeitung 2001, IDS-Korpora)

Beim Romanlesen habe ich einen weiteren entdeckt. Hier hat der Übersetzer ein Erstglied in zwei Varianten realisiert.

*Während der zehn Jahre, in denen sie noch gesund genug war, um weiterhin Gäste zu empfangen, gab Madame Moreau ungefähr ein Abendessen pro Monat. Das erste war ein gelbes Essen: Gougère oder Käsegebäck auf Burgunderart, Hechtklöße holländisch, Wachtelragout mit Safran, Maissalat, Zitronen- und Guavensorbet, dazu Sherry, Château-Chalon, Chateaux-Carbonneux und Eierpunsch mit Sauternes. Das letzte war, 1970, ein schwarzes Essen, serviert auf polierten Schiefertellern; es enthielt natürlich Kaviar, aber auch Tintenfisch auf tarragonische Art, einen **Frischlingrücken** Cumberland, einen Trüffelsalat und Heidelbeer-Charlotte; die Ge-*

¹¹⁰ Übrigens trotz des Aufeinandertreffens der beiden gleichen Konsonanten *f*.

*tränke dieser letzten Mahlzeit waren etwas schwieriger auszuwählen: der Kaviar wurde mit Wodka serviert, der in Basaltbechern ausgeschenkt wurde, und der Tintenfisch mit einem Harzwein von einem in der Tat sehr dunklen Rot, doch für den **Frischlingsrücken** ließ der Butler zwei Flaschen Château-Ducru-Beaucaillou 1955 verabreichen, zu diesem Anlass umgefüllt in Karaffen aus böhmischem Kristall, die genau die erforderliche Schwärze hatten.*

(Perec 2009, S. 536f.)

Und Gisela Zifonun hat mich darauf aufmerksam gemacht, dass sie selbst unabsichtlich manche Erstglieder regelmäßig sogar innerhalb kleinerer Textabschnitte variiert:

*Kann in einer Sprache das **Subjekt-Pronomen** [...] entfallen? Unter welchen Bedingungen geschieht dies? Unter den Vergleichssprachen sind zwei Pro-drop-Sprachen: Das Polnische und das Ungarische. Im Polnischen entfallen [...] **Subjektspronomina**, wenn nicht die Bedingungen für ein ‘starkes’ Pronomenvorkommen gegeben sind.*

(Zifonun 2002, S. 151)

Im Fokus des Linguisteninteresses steht aber die Variation zwischen Individuen. Varianten zwischen Individuen stellen Linguisten seit je bei belletristischen Schriftstellern fest. Variation gehört zum belletristischen Programm. So nennen Fleischer/Barz (1995, S. 138) Zusammensetzungen „wie *Festessucht* (mit -es!) (Ch. Wolf)“. *Festessucht* habe ich im August 2010 tatsächlich nur als Einzelbeleg bei Luukkainen (1997) googeln können, der genau diese Textstelle von Christa Wolf zitiert. In den IDS-Korpora der geschriebenen Sprache kommt es nicht vor. Dort lassen sich aber zahlreiche andere Belege für das Erstglied *Festes* finden, zum Beispiel journalistische und belletristische Belege für *Festesfreude*.

*Die SZENE hat mit dem Fest zur Festspieleröffnung die **Festesfreude** der Salzburger mit einfachen Mitteln geweckt und den Beweis erbracht: Salzburg versteht zu feiern.*

(Salzburger Nachrichten 2000, IDS-Korpora)

*Noch will ich, um der Dankbarkeit willen, einiger gehobener **Festesstunden** in meinem nun schon langen Leben gedenken [...]*

(Thomas Mann 1936, IDS-Korpora)

*Wie denn überhaupt die Aufführung einen Hauch von Intimität und Verinnerlichung gewann, während barocker **Festesjubiläum** – im Werk gewiß auch angelegt – ein wenig in den Hintergrund trat.*

(Die Presse 1999, IDS-Korpora)

*Ob im ORF die Jubiläumssendung, der 400. „Tatort“, „Schwarzer Advent“, am 9. November zu besserer Sendezeit gezeigt wird, war im **Festestaumel** nicht zu eruieren.*

(Kleine Zeitung 1998, IDS-Korpora)

*Die knochentrockene Kälte, die aus der Novelle des Waldviertler Entdeckers des Großen im Kleinen weht, ist gleichfalls bestes literarisches Aspirin gegen die grassierende **Festesbesoffenheit**.*

(Die Presse 1998, IDS-Korpora)

Fleischer/Barz (1995, S. 138) nennen Christa Wolfs *Festessucht* eine „Ausnahme“. Auch *Festestaumel* ist in den IDS-Korpora der geschriebenen Sprache nur 3-mal belegt neben 70 Belegen für *Festtaumel*. *Festesfreude* kommt 80-mal vor neben 446-mal *Festfreude*. Von den 526 Belegen für *Fest* als Erst- und *Freude* als Zweitglied sind also 15,2% Belege für *Festesfreude*. Was ist eine Ausnahme?

Sollten wir Linguisten empfehlen, diese „Ausnahmen“ und überhaupt alle vergleichsweise weniger häufigen Varianten nicht zu verwenden? Sollten wir empfehlen sich an der Mehrheit zu orientieren? „Denn die Menschen suchen Rat in sprachlichen Fragen, und wenn sie sich an das Bewährte halten, können sie zumindest keine größeren Fehler machen. Allerdings wird damit ‘gutes Deutsch’ auf bloßes ‘richtiges Deutsch im Sinne von herkömmlicher Regelinhaltung’ zurückgeführt.“ (Eroms 2007, S. 96).

Auch zeigt uns Schwitalla (2009, S. 234), „dass Variation in einem rhetorischen Sinne ‘nützlich’ ist, unterschiedlichen Rollenanforderungen gerecht zu werden, sich in ein gutes Licht zu setzen, den Adressaten zu überzeugen“, dass sie „auch erfreut“ und dass „sie von den Beteiligten als witzig, überraschend, paradox und möglicherweise entlarvend empfunden wird“. ¹¹¹ *Festessucht* hat in diesem rhetorischen Sinne jedenfalls Fleischer/Barz überrascht; vielleicht fanden sie es witzig.

Auch für Kafka möchte ich individuellen Witz und gutes Licht geltend machen. Nekula (2003, o.S.) wundert sich dagegen, dass Kafka „Wörter benutzt, die es im Hochdeutschen und Binnendeutschen“ in anderen Formen gibt wie „*Fenster***n**Rand, *Werk***s**tage, *Zug***s**Verbindungen, *Ausnahm***e**stag, *Fabriks***c**hef, *Sägen***w**erksbesitzer u.a., andererseits weiß er offensichtlich, dass das Fugenelement -s-, das er durchaus auch dem Standard entsprechend (*Anstaltsfrage*,

¹¹¹ Vgl. auch Schiewe (2007).

Augenblicksbrief, widerspruchslös, Leidenszeit, Professorsfamilie, Gehaltsfrage, Gehaltserhöhung u.a.) bzw. schwankend (*Konzeptbeamten / Konceptsbeamten / Conceptsbeamten, Konzeptsdienst, Rangklasse / Rangklasse / Rangsklasse*) benutzt, als Austriazismus gelten kann, so dass er es nicht nur korrekt (*Hilflosigkeit, Vernunfttheirat, Vernunftgründe, Wohnungsuchen, Aufzugtüre, Mittagessen, Rücksichtnahme*), sondern hyperkorrekt auch da tilgt, wo das Fugenelement hingehört“.

Varianten finden sich aber nicht nur zwischen einzelnen Autoren, auch zwischen Autorengruppen sind Varianten zu beobachten. So haben Ortner et al. (1991, S. 68) ermittelt, dass der Anteil des *s* je nach Textsorte schwankt. Autoren belletristischer Texte verwenden in ihrem Korpus deutlich weniger *s*-Formen als Autoren sonstiger Texte. Dort schwankt zum Beispiel der *s*-Anteil zwischen unter 10% bei Max Frisch und über 30% bei „Wissenschaftstexten“ sowie „Texten der Technik und Presse“.

Wie ist der belletristische Anteil am Gesamt der DeReKo-Schnittstellenanalysen? Noah Bubenhofer hat eine Statistik erstellt, aus der ich in Diagramm 4 die drei häufigsten Erstgliedformen zeige, also die Nullform, die *s*- und die *n*-Form. Ich vergleiche deren prozentuale Anteile beim Gesamt der DeReKo-Schnittstellenanalysen mit den prozentualen Anteilen in einem Teilkorpus aus Texten von Klemperer, den Brüdern Grimm, Goethe und Thomas Mann. Die Statistik bestätigt die Beobachtung von Ortner et al. (1991).

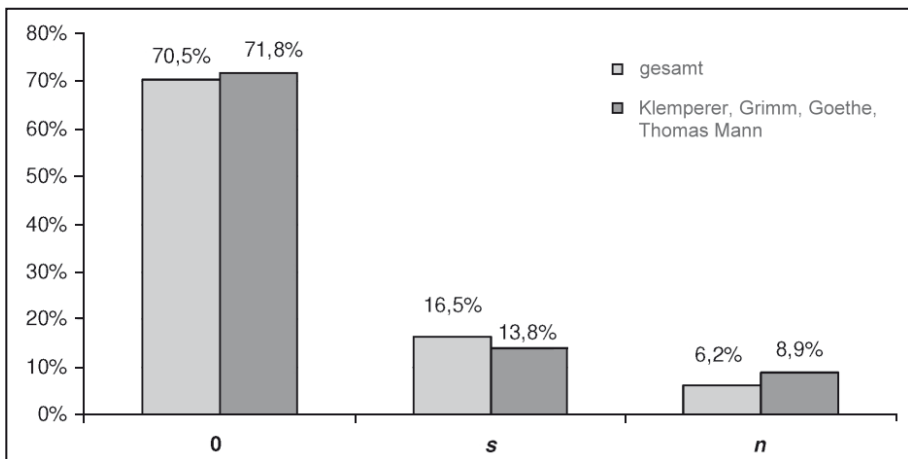


Diagramm 4

Ein Manko der DeReKo-Schnittstellenanalysen ist allerdings, dass Klemperer, die Brüder Grimm, Goethe und Thomas Mann ihre Texte deutlich früher verfasst haben als das Gros der Autoren der DeReKo-Texte. Kann man sie also hinsichtlich des Genres vergleichen? Oder sagt das Diagramm 4 eher etwas aus über heutigen und früheren Sprachgebrauch?

Aufbereitete Teilkorpora aktueller Belletristen stehen nicht zur Verfügung. Hier musste ich auf andere Prüfmöglichkeiten ausweichen. Um rein historische Unterschiede auszuschließen, habe ich die Statistik noch einmal über das Historische Korpus des IDS laufen lassen. Das Historische Korpus enthält 45 Millionen Wörter aus der Fachliteratur, der Belletristik und aus Zeitungen von 1700 bis 1918 und ergibt 75 562 Schnittstellenanalysen. Die Ergebnisse im Diagramm 5 lassen vermuten, dass nicht nur gruppensprachlich belletristische, sondern auch historische Aspekte eine Rolle spielen, denn die Prozente des Historischen Korpus sind, jedenfalls für die Null- und die *s*-Form, denen der vier belletristischen Autoren ähnlich.

Zudem zeigen die Diagramme 4 und Diagramm 5 wiederum nichts anderes als das Vorkommen von Null-, *s*- und *n*-Formen. Ortner et al. (1991, S. 68) vermuten das niedrige Vorkommen belletristischer *s*-Formen im niedrigen Anteil der *s* „nach sich ziehenden“ Abstrakta. Belletristen, meinen Ortner et al. also, verwenden einfach weniger Abstrakta. Um das nachzuprüfen, habe ich typi-

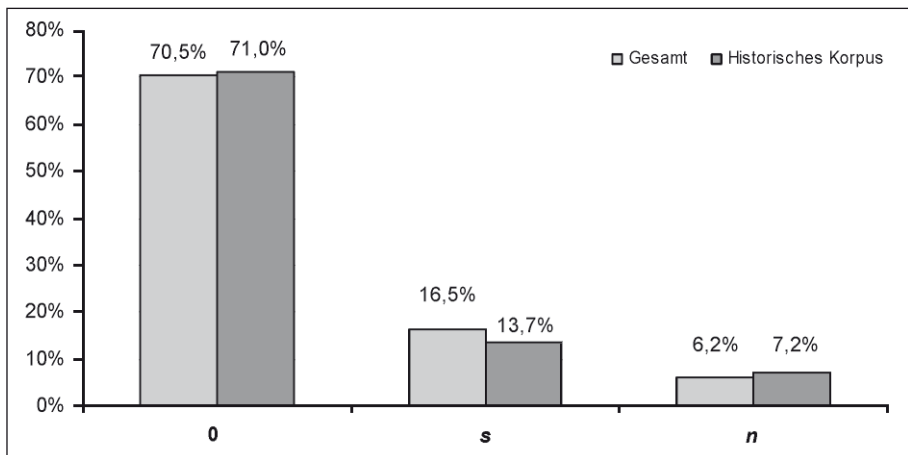


Diagramm 5

sche deutsche Abstrakta recherchiert. Gerechnet habe ich Abstrakta auf *heit*,¹¹² *ion* und *ung*.¹¹³ Tabelle 46 zeigt, wie hoch der Anteil der Erstgliedabstrakta im Vergleich zu den gesamten Erstgliedern in den belletristischen Teilkorpora ist.

	Gesamt	Thomas Mann	Goethe	Grimm
Schnittstellenanalysen	176 405 282	83 634	21 914	5 458
<i>heit</i>	814 924	567	31	0
<i>ion</i>	2 136 744	532	52	0
<i>ung</i>	7 234 838	2 023	705	4

Tabelle 46

Bei den 176 405 282 Schnittstellenanalysen des gesamten DeReKo sind nach Tabelle 46 also zusammen 10 186 506 Erstglieder Abstrakta mit *heit*, *ion* oder *ung*, das sind 5,7% vom Gesamt. Wie sieht das bei den Korpora der belletristischen Autoren aus? Wie das Korpus von Ortner et. al. (1991) zeigen auch die belletristischen DeReKo-Schnittstellenanalysen der Tabelle 47 einen geringeren Abstrakta-Anteil.

	Gesamt	Thomas Mann	Goethe	Grimm
Anteil frequenter Abstrakta	5,7%	3,7%	3,5%	0,07%

Tabelle 47

Verhalten sich andere Autorengruppen anders? Zum Vergleich habe ich Textkorpora verschiedener Autorengruppen recherchiert. Die Korpora ergeben verschieden viele Schnittstellen; zum Vergleich habe ich wieder die Mengen der Schnittstellen angegeben. Die Tabelle 48 zeigt, wieviel Prozent der Schnittstellenanalysen auf die drei häufigsten Schnittstellen entfallen.

¹¹² Nur selten sind *heit*-Substantive keine Abstrakta, etwa *Menschheit*.

¹¹³ Nur selten sind *ung*-Substantive keine Abstrakta, etwa *Waldung*.

	Gesamt	Klemperer Grimm Goethe Mann	FAZ	VDI ¹¹⁴	Lufthansa Bordbuch	Reden Bundestag
Schnittstellenanalysen	176405282	171963	6644614	336978	10772	84692
0	70,5%	71,8%	67,3%	69,2%	78,8%	58,4%
s	16,5%	13,8%	19,0%	18,5%	9,9%	24,2%
n	6,2%	8,9%	5,8%	5,3%	5,7%	4,5%

Tabelle 48

Am meisten weichen die Nullformen vom Gesamt bei den Reden aus dem deutschen Bundestag ab. Dafür sind die *s*-Formen dort deutlich häufiger. Diese Tendenz lässt sich insgesamt feststellen: Wo der Prozentsatz der Nullformen niedriger als im Gesamtkorpus ist, ist der Prozentsatz der *s*-Formen höher. So hat auch die FAZ gegenüber dem Gesamt einen niedrigeren Anteil an Nullformen und einen höheren Anteil an *s*-Formen: Das Gesamt der DeReKo-Schnittstellenanalysen sind zu 16,5% *s*-Formen, die FAZ-Schnittstellenanalysen sind zu 19,0% *s*-Formen. Und auch der Umkehrschluss ist sichtbar: Das Lufthansa Bordbuch hat auffällig viele Nullformen, nämlich 78,8%, und auffällig wenig *s*-Formen, nämlich 9,9%. Vielleicht sagt das gar nichts anderes, als dass in manchen Texten mehr oder weniger Erstglieder mit paradigmatischem oder unparadigmatischem *s* vorkommen;¹¹⁵ vielleicht sagt das aber auch aus, dass Schnittstellen entweder zugunsten der Nullform oder zugunsten der *s*-Form gestaltet werden.

Um die Varianz zwischen Autorengruppen weiter zu prüfen, habe ich für Tabelle 49 vier Erstglieder mit variierenden Schnittstellen recherchiert. Die Schnittstellen variieren die Null- und die *s*-Form: *Abschied* neben *Abschieds*, *Beitrag* neben *Beitrags*, *Quartier* neben *Quartiers* und *Schicksal* neben *Schicksals*. Dabei ergaben sich allerdings mitunter aussagekraftlos geringe Vorkommenszahlen bei den einen versus hohe Vorkommenszahlen bei den anderen: So kommen die Erstgliedvarianten *Beitrag* und *Beitrags* bei Thomas Mann gar nicht vor, dagegen ist *Beitrags* in den VDI-Nachrichten, der FAZ und den

¹¹⁴ VDI Nachrichten – Das Infoportal für Ingenieure.

¹¹⁵ Zur Erinnerung: Die Prozentzahlen sagen etwas aus über die Verwendung von *s*-Formen und nichts über die variante Schnittstellengestaltung.

Bundestagsreden üppig belegt.¹¹⁶ Dafür ist *Schicksals* bei Thomas Mann und in der FAZ gut belegt, in den VDI-Nachrichten und den Bundestagsreden dagegen äußerst schwach. Wirklich hinsichtlich der Gestaltung von Zusammensetzungen vergleichen kann ich also nicht.

	Thomas Mann	FAZ	VDI	Reden Bundestag
<i>Abschied</i>	5	63	0	0
<i>Abschieds</i>	24	1 074	3	4
<i>Beitrag</i>	0	6	0	0
<i>Beitrags</i>	0	2 312	82	144
<i>Quartier</i>	8	48	1	0
<i>Quartiers</i>	0	53	2	1
<i>Schicksal</i>	1	11	0	0
<i>Schicksals</i>	90	470	2	3

Tabelle 49

Neben Autoren belletristischer Texte werden auch Autoren juristischer von Schnittstellenforschern immer mal wieder fokussiert: „Aus der juristischen Fachsprache ist bekannt, dass viele ansonsten obligatorisch zu verfügende Komposita unverfugt bleiben: *Erbschaft+steuer*, *Mehrwert+steuer*, *Schaden+ersatz*“ (Nübling/Szczepaniak 2009, S. 207). Die DeReKo-Schnittstellenanalysen, die vor allem Tages- und Wochenzeitungstexte auswerten, sollten also eher die „ansonsten obligatorischen“ Schnittstellen mit *s* zeigen. Für Tabelle 50 habe ich alle drei Beispiele von Nübling/Szczepaniak recherchiert. Immer tendieren die Schnittstellengestalter zum Unverfugten. Das erstaunt mich selbst, weil ich an *Erbschaftssteuer* und *Schadensersatz* gewöhnt bin. Tatsächlich habe ich im November 2010 auch eine Tendenz zu *Schadensersatz* gegoogelt; dagegen zeigt Google wie die DeReKo-Schnittstellenanalysen eine Tendenz zu *s*-losem *Erbschaftsteuer* und *Mehrwertsteuer*.

¹¹⁶ An den Nullbelegen sieht man übrigens, dass man mit kleinen Korpora spezielle Phänomene oft nicht erfasst.

	DeReKo	Google
<i>Erbschaftssteuer</i>	4 526	192 000
<i>Erbschaftsteuer</i>	5 275	312 000
<i>Mehrwertssteuer</i>	77	70 500
<i>Mehrwertsteuer</i>	33 345	16 900 000
<i>Schadenersatz</i>	7 191	3 450 000
<i>Schadenersatz</i>	18 613	2 100 000

Tabelle 50

Ich vertraue Google. Bei *Schadenersatz* haben Nübling/Szczepaniak (2009) deshalb für mich recht, aber nicht bei *Erbschaft-* und *Mehrwertsteuer*. Wer dennoch daran festhalten will, dass Juristen manche Zusammensetzungen anders gestalten als alle Übrigen, kann natürlich argumentieren, dass die juristischen Termini massiv Eingang in die allgemeine Standardsprache gefunden haben, also auch in Tages- und Wochenzeitungen juristische Gestaltung von Zusammensetzungen nachgeahmt wird. Überprüfen lässt sich das mit automatischen Rechneranalysen natürlich nicht.

Zum Schluss noch ein kleines Lehrstück über Gewohnheiten und Korpusrecherchen. Anfrager fragen uns oft, warum Schnittstellen nicht immer gleich gestaltet werden. Und zwar so, wie der Anfrager es gewohnt ist.

Mail vom 24.2.2004

Sehr geehrte Damen und Herrn,

auf der Suche nach einer Antwort bin ich auf das Institut für deutsche Sprache gestoßen und erhoffe nun von Ihnen Hilfe bei der Klärung meiner Frage: Gibt es eine Regel, nach der ein „s“ zwischen Substantiven eingefügt wird, wenn diese zusammengesetzt werden? Oder anders ausgedrückt: Wieso heißt es Terminsladung aber nicht Terminkalender? Falls Sie selbst die Antwort nicht wissen, können Sie mir bitte sagen, wo ich mich informieren kann?

Ich danke Ihnen herzlich für Ihre Mühe und verbleibe mit freundlichen Grüßen

G.K.

Meine Gewohnheiten sind hier anders. Ich hätte die Frage deshalb so gestellt: Wieso sollte es *Terminkalender* heißen, wenn die Nullform *Termin* vorherrscht? Meine Recherche in den DeReKo-Schnittstellenanalysen zeigt nämlich klare Präferenzen.

Termin	50 866
<i>Terminabsprache</i>	2 893
<i>Terminkalender</i>	14 085
<i>Terminladung</i>	10
Termins	16
<i>Terminsabsprache</i>	6
<i>Terminskalender</i>	0
<i>Terminsladung</i>	0

Tabelle 51

In den IDS-Korpora der geschriebenen Sprache finden sich also keine Belege für *Terminsladung* und auch nur wenige Erstglieder *Termins* mit sonstigen Zweitgliedern.

*Bei der **Terminsabsprache** für das Frühstück im „Club“ im 27. Stock des Atlanta Financial Centers wird klar, daß der Mann begehrt ist wie eh und je.*
(Nürnberger Nachrichten 1996, IDS-Korpora).

Heißt das, dass das Erstglied *Termin* vergleichsweise selten in der Genitivform verwendet wird? Ist *Terminladung* keine übliche Form?

Im September 2010 habe ich rund 5 700 Belege für *Terminsladung* gegoogelt neben 4 700 Belegen für *Terminladung*. Die Googlebelege stammen auch aus Juraforen, Jurablogs und Homepages von Anwaltskanzleien. Heißt das, dass der Anfrager doch zu Recht an *Terminsladung* gewohnt ist, nämlich wie die Mehrheit der Googleschreiber. Heißt das, dass *Terminsladung* Juristengewohnheit ist?

Wie sieht es mit anderen juristischen Gewohnheiten aus? Im September 2010 habe ich rund 27 800 Belege für *Terminsabsprache* gegoogelt neben 447 000 Belegen für *Terminabsprache*. Auch hier sind Juraforen aller Art beteiligt. Heißt das, dass Juristen überwiegend *Terminsladung* sagen, aber *Terminabsprache*?

Die nichtjuristische Zusammensetzung *Terminkalender* habe ich 11 600 000-mal gegoogelt. Die große Zahl der Belege für *Terminskalender* hat mich dann erstaunt: Es fanden sich in Google 5 730 000 Belege für *Terminskalender*. Bei näherem Hinsehen waren das allerdings lauter Belege für *Termin-Kalender*. Google weiß ja oft besser, was ich suche, und ersetzt automatisch mein gesuchtes *s* durch einen Durchkopplungsbindestrich. Heißt das, dass Google

keine brauchbaren Daten liefert?¹¹⁷ Heißt das, dass die DeReKo-Schnittstellenanalysen nicht brauchbar sind? Heißt das, dass Korpusresultate gar nicht so viel sagen, wie wir vielleicht glauben? Denn um nach Krämer (2000, S. 61) mit Friedrich dem Großen zu sprechen: „Mit der ganzen Algebra ist man oftmals nur ein Narr, wenn man nicht noch etwas anderes weiß.“

¹¹⁷ Bickel (2006, o.S.) beantwortet diese Fragen optimistisch, jedenfalls hinsichtlich seiner Recherchen mit AltaVista und Google für das *Variantenwörterbuch des Deutschen* von Ammon et al. (2004). Für ihn „hat sich gezeigt, dass das Internet ein äusserst brauchbares Korpus zur Frequenzabklärung darstellt. Die enorme Grösse und die Vielzahl unterschiedlicher Textsorten, sie reichen von persönlichen Homepages über Verwaltungs- und Gesetzestexte, wissenschaftliche Abhandlungen, Werbung und Dienstleistungsangebote, bis hin zu Zeitungs- und Zeitschriftenarchiven, machen es zu einem äusserst vielseitigen Korpus, das in einem systematischen Bezug zur verschriftlichten Sprachwirklichkeit steht.“

6. Fazit

Eichinger (2009, S. 5) stellt fest, dass „das Zutrauen zur Dokumentation des Sprachgebrauchs als zu befragende Instanz zweifellos zugenommen hat, seit uns hier elektronische Korpora eine gänzlich neue Art der Übersicht erlauben“. Ein Fazit meiner korpusbasierten Studie ist, dass es zu Recht zugenommen hat. Schon deshalb, weil „alles, was man wissen muss, um Erkenntnisse über Sprache zu gewinnen und zu vermitteln, in der Sprache selbst steckt. [...] Das ‘Alles, was man wissen muss’ entnimmt man nicht einzelnen Beispielen oder Fällen, sondern der Spracherwerbende/-lerner/Linguist entnimmt diese als Regelmäßigkeiten, Gesetzmäßigkeiten über Muster aus einem massenhaften Gebrauch von Sprache“ (Perkuhn/Belica 2006, S. 7).

Den massenhaften Gebrauch von Sprache zeigen in meiner Studie die von Noah Bubenhofer erarbeiteten DeReKo-Schnittstellenanalysen mit 176405282 Token-Analysen von authentischen Zusammensetzungen aus Substantiven. Die Analysen waren mein Material zur Prüfung der Regeln für die Wahl von Wortformen und Fugenelementen in Zusammensetzungen. Als Regeln habe ich formuliert: Die Paradigmenregel, die Bedeutungsregel, die Euphonieregel, die Strukturierungsregel und die Gewohnheitsregel.

Alle fünf Regeln finden sich in den DeReKo-Schnittstellenanalysen wieder, konkurrieren allerdings auch. Mal orientieren sich Sprecherschreiber an der einen, mal an der anderen Regel. Das finden viele Lerner und Linguisten unbefriedigend. Allgemein gilt ja, dass „wir Menschen ein starkes Verlangen nach Regeln haben, weil wir die Dimensionen der Dinge so reduzieren müssen, dass sie in unseren Kopf passen. Besser gesagt, bedauerlicherweise: dass wir sie in unseren Kopf quetschen können. [...] Interessanterweise besteht genau dieses Verlangen nach Ordnung auch bei wissenschaftlichen Unterfangen; der offizielle Zweck der Wissenschaft ist allerdings im Gegensatz zur Kunst, die Wahrheit herauszufinden, nicht, uns ein Gefühl der Organisation zu geben oder dazu zu führen, dass wir uns besser fühlen“ (Taleb 2010, S. 95f.).

Meine wissenschaftliche Wahrheit ist, dass die Gestaltungsmöglichkeiten deutscher Zusammensetzungen nicht wirklich in unseren Kopf passen. Jedenfalls nicht so, dass wir Linguisten zuverlässig voraussagen könnten, wie Zusammensetzungen im Einzelnen organisiert werden. Wir können aber immerhin einen Überblick geben, Tendenzen aufzeigen und das Durcheinander der Details beschreiben.

Beschrieben habe ich unter anderem, wie ähnliche Erstglieder mal in dieser, mal in anderer Form gewählt werden: Zum Beispiel sagen wir *Hundehütte*, aber nicht *Schafestall*. Auch identische Erstglieder variieren. In den DeReKo-Schnittstellenanalysen sind rund 20% der identischen Erstglieder variant. Die meisten kommen in zwei Varianten vor: *Arzt* neben *Ärzte*, manche auch in drei Varianten: *Betrieb* neben *Betriebe* neben *Betriebs*. Erstglieder in mehr als drei Varianten sind selten: *Kind* neben *Kinds* neben *Kindes* neben *Kinder*.

Einzelne variante Erstglieder zeigen meist eine Tendenz zu der einen oder anderen Variante. Von dieser Tendenz zu schließen auf Tendenzen bei ähnlichen Erstgliedern, führt jedoch schnell irre. So gibt es zum Beispiel beim Erstglied *Generation* eine Tendenz zur Variante *Generationen*, beim Erstglied *Aktion* dagegen gibt es keine Tendenz zu *Aktionen*, sondern eine zu *Aktions*. Das illustriert, worauf schon Lüttke/Mattheier (2005) hinweisen:

Die Erkenntnis, dass wir [...] mit dem, was die deskriptive Grammatik eine Sprache nennt, mit der Zusammenfassung des Usuellen, überhaupt gar nicht rechnen dürfen als einer 'Abstraktion, die keine reale Existenz hat', wie Hermann Paul betont, ist im Laufe unseres Jahrhunderts, auch in Folge Saussures, verloren gegangen und muss immer wieder neu entdeckt werden. Zwar weiß man, dass eine Sprache nicht einheitlich ist, jedoch legt man dieses Wissen der Forschung nicht stets und grundsätzlich zugrunde. (Lüttke/Mattheier 2005, S. 24).

Die Varianz der Gestaltung von Zusammensetzungen gibt nun auch Anlass, neu zu überlegen, was wir unter Standard verstehen wollen und ob uns die Kategorisierung in Standard und Nichtstandard überhaupt weiter bringt.¹¹⁸

Variatio delectat gilt auch für die besonders unter Lernern verrufene Gestaltung deutscher Zusammensetzungen. *Variante* ist entlehnt aus französisch *variante*, das meistens mit 'Abweichung' übersetzt wird, das ich aber kreativ-freundlicher mit 'Lesart, Spielart' übersetzen möchte. Deshalb gebe ich zwar einerseits Klein (2003) irgendwie recht:

In der Existenz von Zweifelsfällen verkörpert sich die Tatsache, dass in einer Sprachgemeinschaft ein Orientierungsbedürfnis in sprachlichen Dingen besteht. Es gehört zu den Aufgaben der Sprachwissenschaft, sich mit diesem Orientierungsbedürfnis auseinanderzusetzen, indem sie auf deskriptiver Basis normativ interpretierbare Anhaltspunkte zur Klärung der Zweifelsfälle identifiziert. (Klein 2003, o.S.).

¹¹⁸ Im Rahmen des Projekts „Grammatische Variation im Deutschen“ hat dazu im Sommer 2010 ein Workshop stattgefunden, dessen Ergebnisse voraussichtlich publiziert werden.

Andererseits fragen wir uns mit Eichinger (2005, S. 368) „Normen ja – aber wem seine?“¹¹⁹ Andererseits befreit Varianz aber auch: Wenn alles Mögliche möglich ist, kann jeder von uns weniger falsch machen.

Schließlich habe ich aus meiner Studie gelernt, dass wir Linguisten uns ununterbrochen fragen sollten, wie es mit unserer Sprachkompetenz aussieht. Ich jedenfalls war während der Erarbeitung vieler Tabellen oft verblüfft über den massenhaften, mitunter dominanten Gebrauch einer Variante, die ich bis dahin für marginal und auffällig gehalten hatte. Falls wir Linguisten also normativ Interpretierbares formulieren wollen, sollten wir noch genauer hinsehen. Normen ja – aber woher?

¹¹⁹ Vgl. zu den Instanzen der Sprachnormierung präzise Ammon (2005).

7. Literatur

- Ammon, Ulrich (2005): Standard und Variation: Norm, Autorität, Legitimation. In: Eichinger/Kallmeyer (Hg.), S. 28-40.
- Ammon, Ulrich/Bickel, Hans/Ebner, Jakob/Esterhammer, Ruth/Gasser, Markus/Hofer, Lorenz/Kellermeier-Rehbein, Birte/Löffler, Heinrich/Mangott, Doris/Moser, Hans/Schläpfer, Robert/Schloßmacher, Michael/Schmidlin, Regula/Vallaster, Günter unt. Mitarb. v. Kyvelos, Rhea/Nyffenegger, Regula/Oehler, Thomas (2004): Variantenwörterbuch des Deutschen – Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol. Berlin.
- Augst, Gerhard (1975): Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache. (= Forschungsberichte des Instituts für Deutsche Sprache 25). Tübingen.
- Augst, Gerhard (2001): Gefahr durch lange und kurze Wörter? Lang- und Kurzwortgefahr? LKW-Gefahr? In: Stickel, Gerhard (Hg.): Neues und Fremdes im deutschen Wortschatz – Aktueller lexikalischer Wandel. Berlin/New York, S. 210-238.
- Barz, Irmhild (2005): Die Wortbildung. In: Duden – Die Grammatik. Unentbehrlich für richtiges Deutsch. 7., völl. neu bearb. u. erw. Aufl. Hrsg. v. der Dudenredaktion. (= Duden 4). Mannheim u.a., S. 641-772.
- Bickel, Hans (2006): Das Internet als linguistisches Korpus. In: Linguistik online 28, 3, o. S. http://www.linguistik-online.de/28_06/bickel.html (Stand: 8/2011).
- Bourdain, Anthony (2007): Kleine Schweinereien – serviert vom Autor des Bestsellers „Geständnisse eines Küchenchefs“. Aus dem Amerikanischen von Ronald Gutlebert. München.
- Burkhardt, Armin (Hg.) (2007): Was ist gutes Deutsch? Studien und Meinungen zum gepflegten Sprachgebrauch. (= Thema Deutsch 8). Mannheim/Leipzig/Wien/Zürich.
- Christen, Helen (2004): Vorwort: Vom Wissen um die soziale Komponente arealer Sprachvariation zu ihrer Erforschung. In: Christen, Helen unt. Mitarb. v. Agnès Noyer (Hg.): Dialekt, Regiolekt und Standardsprache im sozialen und zeitlichen Raum. Beiträge zum 1. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen, Marburg/ Lahn, 5.-8. März 2003. Wien, S. 7-20.
- Cornelissen, Georg (2005): Rheinisches Deutsch – Wer spricht wie mit wem und warum. Köln.
- Curtin, Michael (2009): Der Club der Weihnachtshasser. Roman. Deutsch von Friederike Zeininger. München.

- Darski, Józef (1996): Die Fugenelemente im Deutschen. In: Sroka, Kazimierz A. (Hg.): *Kognitive Aspekte der Sprache: Akten des 30. Linguistischen Kolloquiums, Gdańsk 1995*. Tübingen, S. 49-54.
- Deutsches Wörterbuch von Jacob und Wilhelm Grimm (1854 [1984]). Fotomech. Nachdr. Bd. 1. München.
- Deutsches Wörterbuch von Jacob und Wilhelm Grimm (1878 [1984]). Fotomech. Nachdr. Bd. 4. München.
- Deutsches Wörterbuch von Jacob und Wilhelm Grimm (1899 [1984]). Fotomech. Nachdr. Bd. 15. München.
- Deutsches Wörterbuch von Jacob und Wilhelm Grimm (1941 [1984]). Fotomech. Nachdr. Bd. 18. München.
- Donalies, Elke (1999): Präfixverben, Halbpräfixverben, Partikelverben, Konstitutionsverben oder verbale Gefüge? Ein Analyseproblem der deutschen Wortbildung. In: *Studia Germanica Universitatis Vesprimiensis* 3, S. 127-143.
- Donalies, Elke (2007): *Basiswissen Deutsche Wortbildung*. (= UTB 2876). Tübingen/Basel.
- Dürscheid, Christa (2006): *Einführung in die Schriftlinguistik*. 3., überarb. u. erg. Aufl. Ergänzt um ein Kapitel zur Typographie von Jürgen Spitzmüller. (= Studienbücher zur Linguistik 8). Göttingen.
- Eichinger, Ludwig M. (2000): *Deutsche Wortbildung. Eine Einführung*. (= narr studienbücher). Tübingen.
- Eichinger, Ludwig M. (2004): Von Köpfen, Nägeln und anderen guten Bekannten. In: Steyer, Kathrin (Hg.): *Wortverbindungen – mehr oder weniger fest. Jahrbuch 2003 des Instituts für Deutsche Sprache*. Berlin/New York, S. 1-18.
- Eichinger, Ludwig M. (2005): Standardnorm, Sprachkultur und die Veränderung der normativen Erwartungen. In: Eichinger/Kallmeyer (Hg.), S. 363-381.
- Eichinger, Ludwig M. (2009): Begrüßung: Vom grammatischen Wissen und seiner vernünftigen Verwendung. In: Konopka/Strecker (Hg.), S. 1-6.
- Eichinger, Ludwig M./Kallmeyer, Werner (Hg.) (2005): *Standardvariation – Wie viel Variation trägt die deutsche Sprache? Jahrbuch 2004 des Instituts für Deutsche Sprache*. Berlin/New York.
- Elshorbagy, Fathy (2007): *Fugenelemente der deutschen Komposita*. Univ.-Diss. Düsseldorf.
- Eroms, Hans-Werner (2007): Grammatisch gutes Deutsch – mehr als nur richtiges Deutsch. In: Burkhardt (Hg.), S. 90-108.

- Eschenlohr, Stefanie (1999): Vom Nomen zum Verb: Konversion, Präfigierung und Rückbildung im Deutschen. (= Germanistische Linguistik Monographien 3). Hildesheim u.a.
- Fehringer, Carol (2009): *Wie wär's mit einem richtigen Mädelsabend?* Plural-s within Compounds in Colloquial Northern German. In: Journal of Germanic Linguistics 21, 2, S. 149-165.
- Fehringer, Carol (i.Dr.): Nominale Diminutiva bei Komposita im Schwäbischen: Ein „Words and Rules“-Ansatz. In: Zeitschrift für Dialektologie und Linguistik.
- Figueras, Marcelo (2010): Das Lied von Leben und Tod. Roman. Aus dem Spanischen von Sabine Gierberg. München.
- Fleischer, Wolfgang/Barz, Irmhild unt. Mitarb. v. Marianne Schröder (1995): Wortbildung der deutschen Gegenwartssprache. 2., durchges. u. erg. Aufl. Tübingen.
- Fuhrhop, Nanna (1998): Grenzfälle morphologischer Einheiten. (= Studien zur deutschen Grammatik 57). Tübingen.
- Fuhrhop, Nanna (2000): Zeigen Fugenelemente die Morphologisierung von Komposita an? In: Thieroff, Rolf/Tamrat, Matthias/Fuhrhop, Nanna/Teuber, Oliver (Hg.): Deutsche Grammatik in Theorie und Praxis. Tübingen, S. 201-213.
- Hartkamp, Sandra/Schneider-Wiejowski, Karina (2010): Die Regelkonformität in der Distribution von Fugenelementen – Eine korpuslinguistische Untersuchung neologistischer Substantivkomposita mit nominalem Erstglied. In: Muttersprache 120, S. 198-213.
- Heringer, Hans Jürgen (1984): Wortbildung: Sinn aus dem Chaos. In: Deutsche Sprache 12, S. 1-13.
- Heringer, Hans Jürgen (2009a): Ist das Deutsche grammatisch zu fassen? In: Konopka/Strecker (Hg.), S. 23-32.
- Heringer, Hans Jürgen (2009b): Morphologie. (= Libac – Linguistik für Bachelor). Paderborn.
- Hyvärinen, Irma (1997): Wortbildung, Textanalyse und Stil: Bindestrichkomposita im deutsch-finnischen Kontrast. In: Fix, Ulla (Hg.): Stile, Stilprägungen, Stilgeschichte. Über Epochen-, Gattungs- und Autorenstile. Sprachliche Analysen und didaktische Aspekte. (= Sprache – Literatur und Geschichte. Studien zur Linguistik/ Germanistik 15). Heidelberg, S. 281-324.
- Janich, Nina (2004): Die bewusste Entscheidung – Eine handlungsorientierte Theorie der Sprachkultur. Tübingen.
- Jean Paul (1819): Über das Zusammenfügen der deutschen Doppelwörter. In: Miller, Norbert/Schmidt-Biggemann, Wilhelm (Hg.) (1978): Jean Paul – Sämtliche Werke, 2. Abteilung: Jugendwerke und Vermischte Schriften. München.

- Kauffer, Maurice (1993): La composition nominale en allemand – Etude linguistique et statistique des cartes de restaurants. (= Göppinger Arbeiten zur Germanistik 568). Göppingen.
- Kellermeier-Rehbein, Birte (2005): Areale Wortbildungsvarianten des Standarddeutschen – Beiuntersuchung zum Variantenwörterbuch des Deutschen. (= Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 61). Frankfurt a.M. u.a.
- Kilgarrriff, Adam (2006): Googleology is Bad Science. In: Computational Linguistics 33, S. 147-151.
- Klein, Wolf Peter (2003): Sprachliche Zweifelsfälle als linguistischer Gegenstand. Zur Einführung in ein vergessenes Thema der Sprachwissenschaft. In: Linguistik online 16, 4, o.S. http://www.linguistik-online.de/16_03/klein.html (Stand: 8/2011).
- Konopka, Marek/Strecker, Bruno (Hg.) (2009): Deutsche Grammatik – Regeln, Normen, Sprachgebrauch. Jahrbuch 2008 des Instituts für Deutsche Sprache. Berlin/New York.
- Krämer, Walter (2000): So lügt man mit Statistik. München/Zürich.
- Krott, Andrea/Schreuder, Robert/Baayen, R. Harald/Dressler, Wolfgang U. (2007): Analogical effects on linking elements in German compounds. In: Language and Cognitive Processes 22, S. 25-57.
- Kürschner, Sebastian (2003): Von *Volk-s-musik* und *Sport-Ø-geist* im *Lemming-Ø-land* – af *folk-e-musik* og *sport-s-ånd* i *lemming-e-lande*: Fugenelemente im Deutschen und im Dänischen – eine kontrastive Studie zu einem Grenzfall der Morphologie. Mag.-Arb. Univ. Freiburg. http://www.freidok.uni-freiburg.de/volltexte/1256/pdf/Kuerschner_Fugenelemente_MagArbeit.pdf (Stand: 8/2011).
- Lauterbach, Stefan (1993): Genitiv, Komposition und Präpositionalattribut – Zum System nominaler Relationen im Deutschen. München.
- Lenz, Alexandra N./Mattheier, Klaus J. (Hg.) (2005): Varietäten – Theorie und Empirie. (= VarioLingua 23). Frankfurt a.M. u.a.
- Lüdtke, Jens/Mattheier, Klaus J. (2005): Variation – Varietäten – Standardsprachen. Wege für die Forschung. In: Lenz/Mattheier (Hg.), S. 13-38.
- Luukkainen, Matti (1997): These, Antithese, Synthese: Zu Wandel und Beständigkeit des Sprachstils im Werk von Christa Wolf 1961-1996. Hamburg.
- Marcus, Gary/Brinkmann, Ursula/Clahsen, Harald/Wiese, Richard/Pinker, Steven (1995): German inflection: The exception that proves the rule. In: Cognitive Psychology 29, S. 189-256.
- Mark Twain (1876 [1999]): The Awful German Language – Die schreckliche deutsche Sprache. 4. Aufl. Waltrop/Leipzig.

- Michel, Sascha (2009): *Schaden-0-ersatz* vs. *Schaden-s-ersatz*. Ein Erklärungsansatz synchroner Schwankungsfälle bei der Fugenbildung von N+N-Komposita. In: Deutsche Sprache 37, S. 334-351.
- Michel, Sascha (2010): Zur Systematik der Kompositionsstammformbildung bei N+N-Komposita – Implikationen für den DaF-Unterricht. Ms. (angenommen).
- Nekula, Marek (2003): Franz Kafkas Deutsch. In: Linguistik online 13, 1, o.S. http://www.linguistik-online.de/13_01/nekula.html (Stand: 8/2011).
- Nübling, Damaris / Szczepaniak, Renata (2008): On the way from morphology to phonology: German linking elements and the role of phonological word. In: Morphology 18, 1-25.
- Nübling, Damaris / Szczepaniak, Renata (2009): *Religion+s+freiheit, Stabilität+s+pakt* und *Subjekt(+s+)pronomen*: Fugenelemente als Marker phonologischer Wortgrenzen. In: Müller, Peter O. (Hg.): Studien zur Fremdwortbildung. (= Germanistische Linguistik 197-198). Hildesheim u.a., S. 195-222.
- Ortner, Lorelies / Müller-Bollhagen, Elgin / Ortner, Hans-Peter / Wellmann, Hans / Pümpel-Mader, Maria / Gärtner, Hildegard (1991): Deutsche Wortbildung. Typen und Tendenzen in der Gegenwart. Eine Bestandsaufnahme des Instituts für Deutsche Sprache, Forschungsstelle Innsbruck. Vierter Hauptteil: Substantivkomposita. (= Sprache der Gegenwart LXXIX). Berlin / New York.
- Pavlov, Vladimir M. (2004): Zur Entwicklung der substantivischen Zusammensetzung im Frühneuhochdeutschen. In: Mattheier, Klaus J. / Nitta, Haruo (Hg.): Sprachwandel und Gesellschaftswandel – Wurzeln des heutigen Deutsch. Studien des deutsch-japanischen Arbeitskreises für Frühneuhochdeutschforschung. München, S. 99-119.
- Perkuhn, Rainer / Belica, Cyril (2006): Korpuslinguistik – Das unbekannte Wesen oder Mythen über Korpus und Korpuslinguistik. In: Sprachreport 22, 4, S. 2-8.
- Plank, Frans (1981): Morphologische (Ir-)Regularitäten. Aspekte der Wortstrukturtheorie. (= Studien zur deutschen Grammatik 13). Tübingen.
- Paul, Hermann (1975): Prinzipien der Sprachgeschichte. 9., unveränd. Aufl., Studienausg. Tübingen.
- Perec, Georges (2009): Das Leben – Gebrauchsanweisung. Romane. Deutsch von Eugen Helmlé. Frankfurt a.M.
- Ramers, Karl Heinz (1997): Die Kunst der Fuge: Zum morphologischen Status von Verbindungselementen in Nominalkomposita. In: Dürscheid, Christa / Ramers, Karl Heinz / Schwarz, Monika (Hg.): Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag. Tübingen, S. 33-46.
- Ribeiro, Tarcísia Múcia Otte (1996): Bemerkungen zu Semantisierungsmöglichkeiten der Fugenelemente. In: Zielsprache Deutsch 27, S. 14-20.

- Ruge, Nikolaus (2006): Methodische Bemerkungen zur Untersuchung der so genannten Bindestrichkomposita Thomas Manns. In: *Wirkendes Wort* 56, S. 475-488.
- Satkauskaitė, Danguolė (2008): Die Rolle des Bindestrichs bei der Konventionalisierung von Zusammensetzungen. In: *Linguistics – Germanic and Romance Studies* 59, S. 253-260.
- Schiewe, Jürgen (2007): Angemessenheit, Prägnanz, Variation. Anmerkungen zum guten Deutsch aus sprachkritischer Sicht. In: Burkhardt (Hg.), S. 369-380.
- Schmidt, Jürgen Erich (2005): Versuch zum Varietätenbegriff. In: Lenz/Mattheier (Hg.), S. 61-74.
- Schwitalla, Johannes (2009): Sprachvariation als Ressource für unterschiedliche Stimmen des Ichs. In: Gilles, Peter/Scharloth, Joachim/Ziegler, Evelyn (Hg.): *Variatio delectat – Empirische Evidenzen und theoretische Passungen sprachlicher Variation*. Festschrift für Klaus J. Mattheier zum 65. Geburtstag. (= *VarioLingua* 37). Frankfurt a.M. u.a., S. 215-238.
- Spiekermann, Helmut (2008): Sprache in Baden-Württemberg – Merkmale des regionalen Standards. (= *Linguistische Arbeiten* 526). Tübingen.
- Szagan, Gisela (2006): Sprachentwicklung beim Kind. Völl. überarb. Neuausg. Weinheim/Basel.
- Taleb, Nassim Nicholas (2010): Der Schwarze Schwan – Die Macht höchst unwahrscheinlicher Ereignisse. Aus dem Amerikanischen von Ingrid Proß-Gill. Frankfurt a.M./Wien/Zürich.
- Toutonghi, Pauls (2009): Die Geschichte von Yuri Balodis und seinem Vater, der eigentlich Country-Star war. Roman. Übersetzt von Eva Bonné. Berlin.
- Valentin, Karl (2006): Mein komisches Wörterbuch – Sprüche für alle Lebenslagen. München.
- Vargas, Fred (2009): Vom Sinn des Lebens, der Liebe und dem Aufräumen von Schränken. Aus dem Französischen von Christel Gersch. Berlin.
- Varsami, John Alex (2007): Fugenelemente im Deutschen. Hausarb. Univ. Stuttgart.
- Wall, Larry/Christiansen, Tom/Orwant, Jon (2001): Programmieren mit Perl. Köln u.a.
- Wegener, Heide (2003): Entstehung und Funktion der Fugenelemente im Deutschen, oder: Warum wir keine **Autosbahn* haben. In: *Linguistische Berichte* 196, S. 245-257.
- Wegener, Heide (2005): Das Hühnerei vor der Hundehütte. Von der Notwendigkeit historischen Wissens in der Grammatikographie des Deutschen. In: Berner, Elisabeth et al. (Hg.): *Ein gross und narhafft haffen*. Festschrift für Joachim Gessinger. Potsdam, S. 175-187.

- Willems, Klaas (2001): Produktivität, syntaktische Struktur und Norm. Deskriptive Normregularitäten transparenter nominaler Wortbildungsmuster und kontrastive Wortbildungsforschung. In: *Zeitschrift für Germanistische Linguistik* 29, S. 143-166.
- Yglesias, Rafael (2002): *Dr. Nerudas Therapie gegen das Böse. Roman. Aus dem Amerikanischen von Kurt Neff.* Frankfurt a.M.
- Zifonun, Gisela (2002): Neue Wege in der vergleichenden Grammatikschreibung. In: Ágel, Vilmos/Herzog, Andreas (Hg.): *Jahrbuch der ungarischen Germanistik* 2001. Budapest, S. 143-155.
- Zifonun, Gisela (2010): Possessive Attribute im Deutschen. In: *Deutsche Sprache* 38, S. 124-153.